

References

- Cooreman, A. 1994. A Functional Typology of Antipassives. In: Fox, B. A., Hopper, P. J. (ed.): *Voice: Form and Function*. Amsterdam–Philadelphia, 49-88.
- Creissels, D. 2006. *Syntaxe générale: une introduction typologique 1: Catégories et constructions*. Paris.
- Foley, W. A., Van Valin, R. D. 1985. *Functional syntax and universal grammar*. Cambridge–New York.
- Genusiene, E. 1987. *The Typology of Reflexives*. Berlin–New York–Amsterdam. (Coll. Empirical Approaches to Language Typology 2).
- Gerritsen, N. 1999. *Russian Reflexive Verbs: Studies in Slavic and General Linguistics, 15*. Amsterdam–Atlanta.
- Haspelmath, M., Muller-Bardey, T. 2000. Valence change. In: Booij, G., Lehmann, C., Mugdan, J. (ed.): *Morphology: A Handbook on Inflection and Word Formation*. Berlin, 1-23.
- Haspelmath, M, et al. (ed.). 2005. *The World Atlas Of Language Structures*. Oxford.
- Kretov, A. A. 1978. Osobennosti semantiki vozvratnych glagolov vključenno go neoduševlennogo ob'jekta v russkom jazyke. In: Sobinnikova, V.I. (Red.): *Materialy po russko-slavjanskomu jazykoznaniju*. Voronež, 162-168.
- Lazard, G., Moysse-Faurie, C. (ed.). 2005. *La Linguistique typologique*. Villeneuve d'Ascq. (Coll. Sens et Structures).
- Levin, B. 1993. *English Verb Classes and Alternations. A Preliminary Investigations*. Chicago.
- Nouguier-Voisi, S. 2005. Antipassif et langues accusatives. In: Lazard, G., Moysse-Faurie, C. (ed.): *Linguistique typologique*. Villeneuve d'Ascq, 193-205.
- Polinsky, M. 2005. Antipassive Constructions. In: Haspelmath, M, et al. (ed.). *The World Atlas Of Language Structures*. Oxford, 438-439.
- Say, S. 2005. The pragmatic motivation of antipassive in Russian. In: Cap, P. (ed.). *Pragmatics today*. Frankfurt/Main, 421-440.
- Say, S. 2005. Antipassive sja-verbs in Russian: Between inflection and derivation. In: Dressler, U. W., Kastovsky, D., Pfeiffer, O. et al. (ed.). *Morphology and its demarcations: selected papers from the 11th Morphology Meeting, Vienna, February 2004*. Amsterdam–Philadelphia, 253-275.
- Tesnière, L. 1959. *Elément de syntaxe structurale*. Paris.
- Whaley, L. 1997. *Introduction to Typology: The Unity and Diversity of Language*. Thousand Oaks.

Streszczenie

Celem tego artykułu jest zbadanie zależności między zmianą formy morfologicznej predykatu a jego walencją syntaktyczną. Analiza wykonana jest w oparciu o konstrukcję antypasywną w języku polskim i rosyjskim.

Lyon

(katarzyna.janic@univ-lyon2.fr)

Katarzyna Janic

ZUM ANALYTISMUS UND SYNTHETISMUS IN
SLAWISCHEN SPRACHEN: MORPHOLOGISCHE
WORTSTRUKTUR IN PARALLELTEXTEN

0. Einleitung

Eine bekannte morphologische Klassifikation von Sprachen ist die Unterscheidung zwischen einem analytischen und einem synthetischen Typus. Diese Unterscheidung geht auf die Anfänge sprachtypologischer Überlegungen zurück und lässt sich mit bekannten Namen wie den Brüdern Schlegel, Wilhelm von Humboldt und Georg von Gabelentz in Verbindung bringen (vgl. Bossong 2004, Hinrichs 2004). Vereinfacht gesagt geht es bei dieser Unterscheidung von Sprachen um zwei unterschiedliche Arten der morphologischen und morphosyntaktischen Organisation, die zwischen zwei Polen anzusiedeln ist: Zwischen morphologisch „formlosen“ (d.h. Sprachen u.a. ohne Nominalflexion) und morphologisch „geformten“ Sprachen. An dieser einfachen morphologischen Charakterisierung von Sprachen fand die Sprachwissenschaft kein Genügen, so dass in weiterer Folge eine Spezifizierung erfolgte und neben dem analytischen und synthetischen ein isolierender, flektierender, agglutinierender, inkorporierender usw. Typus unterschieden wurde. Dabei wird allerdings nur innerhalb der flektierenden Sprachen zwischen einem analytischen und einem synthetischen Typus unterschieden. Zu Details vgl. den kenntnisreichen wissenschaftshistorischen Abriss in Coseriu (1972).

In letzter Zeit ist für die slawistische Linguistik in einem gewissen Sinne eine Neu-entdeckung der Begriffe „analytisch“ / „synthetisch“ auszumachen. Der Versuch einer Synthese von unterschiedlichen Ansätzen und der einer Neu-Interpretation der beiden Begriffe geht auf Hinrichs (2004a:21) zurück. Der verbindet mit dem Begriff „analytisch“ die Vorstellung, dass lexikalische vs. grammatische Informationen von der morphologischen auf die syntaktische Ebene verlagert werden. Dies bedeutet, dass grammatische Information tendenziell außerhalb eines Volllexems, d.h. also „autonom“, ausgedrückt wird. Oder anders: Bei synthetischen Sprachen tritt eine Mehrkomponenten-Konstruktion auf, die potentiell durch einfache flektivische Konstruktionen ausgedrückt werden könnte (Reiter 2004:87).

Mit dieser Spezifizierung der Begriffe „analytisch“ / „synthetisch“ sind aber auch eine Reihe von Folgeproblemen verbunden, wie etwa die Fragen nach der Wort-Einheit und der Wortgrenze, die jedoch aus heutiger Sicht kaum in übersprachlicher Weise gelöst werden können (Hinrichs 2004:21). Eine radikale Konsequenz aus der fehlenden universal gültigen Bestimmung der Begriffe „Wort“ bzw. „Wortgrenzen“ ist, wie in Schwegler (1990:46) ausgeführt, der völlige Verzicht auf eine Klassifikation in analytische und synthetische Sprachen. Ähnlich äußert sich auch Haarmann (2004:73), der meint, dass die absolute Gültigkeit von Wortgrenzen ein theoretisches Konstrukt sei (was im Übrigen für alle sprachlichen Konstrukte gilt). Vor allem aber verweist er auf eine Überbewertung des Wortbegriffes, da in vielen Sprachen eine satzphonetische Verkettung Priorität vor einer Einzelwortlautung hat. Des Weiteren ist zu bedenken, dass je nach angenommener Wortdefinition die Analyse bestimmter Phänomene verschiedenartige Ergebnisse zeigt.

Die Diskussion um die Möglichkeiten einer stringenten Wortdefinition (analog dazu auch die Definition von „Wortform“, „Lexem“ usw.) kann in der Linguistik auf eine sehr lange Tradition zurückblicken. Vgl. hierzu die Übersicht in Krámský (1969) über eine Vielzahl von unterschiedlichen Wortdefinitionen. Auch neuere Auseinandersetzungen

gen (vgl. dazu u.a. den Sammelband von Dixon, Aikhenvald 2002a) mit dieser Problematik gehen davon aus, dass kaum eine allgemeingültige und sprachenübergreifende Definition dieses für die Linguistik zentralen Begriffes möglich scheint. Allerdings kristallisieren sich durch den Bezug auf phonetisch-phonologische, morphologische, grammatikalische und syntaktische Kriterien zumindest Umrisse eines Wortbegriffes heraus, der für weiterführende linguistische Analysen operationalisiert werden kann (vgl. dazu Dixon, Aikhenvald 2002b:6ff.).

In Bezug auf slawische Sprachen scheint aber keine Notwendigkeit zu bestehen, auf den Wortbegriff und somit auf eine morphologische Klassifikation der Sprachen entlang der beiden Pole „analytisch“ und „synthetisch“ zu verzichten. Der folgende Beitrag stellt sich nun die Aufgabe, 1. eine quantitativ messbare Eigenschaft der morphologischen Struktur von slawischen Sprachen, nämlich die Wortlänge gemessen in der Anzahl von Morphemen, als geeignete Eigenschaft für die Bestimmung des Grades an Analytizismus zu diskutieren und 2. bei der vergleichenden Untersuchung der Wortlänge auf eine möglichst homogene Datenbasis in Form von Paralleltexten (Russisch, Slowakisch, Slowenisch, Mazedonisch, Bulgarisch) zurückzugreifen.

1. Linguistische Merkmale und Eigenschaften

Es sind an unterschiedlichen Stellen Vorschläge zu finden, auf welche Art und Weise die Analytizität / Synthetizität in Sprachen untersucht werden kann. In Bezug auf slawische Sprachen werden u.a. folgende linguistische Merkmale (vgl. Hinrichs 2000a:92 und 2004:19ff.) genannt: (1) Kasusverfall (unter Einschluss von Kasuslabialität, Kasusverlust, Vermengung und Austausch von Orts- und Richtungskasus), Kasussynekretismus, Kasusreichtum; (2) Durchbrechen von Kongruenzen bzw. Abbau von Kongruenzbeziehungen; (3) Desemantisierung und Produktivität von Präpositionen, Stärkung von Präpositionen; (4) Verwendung von analytischen / synthetischen Komparativen; (5) periphrastisches Futur, Infinitiv-Ersatz, allgemein die Multiverbierungen von z.B. Funktionsverbgefügen; (6) allgemeiner Ausdrucksdruck von grammatischen Beziehungen innerhalb oder außerhalb des Wortes; (7) Präfigierungserscheinungen beim Verb und (8) defektive Erscheinungen im Rahmen der Wortbildung. Damit sind sicherlich nicht alle Untersuchungsmöglichkeiten des analytischen und synthetischen Sprachbaus ausgeschöpft, aber man bekommt eine ausreichende Vorstellung über die Bandbreite des Phänomens. Es wird im Folgenden auch keine Erweiterung dieser Bandbreite angestrebt, sondern es wird vielmehr – ein Desiderat von Hinrichs (2000a:84-85) aufnehmend – der Grad der Analytizität in einzelnen Sprachen bestimmt, um so einen Vergleich zwischen mehreren Sprachen durchführen zu können.

2. Versuche der Messung von Analytizität und Synthetizität

Abgesehen von diesen vielfältigen linguistischen Untersuchungsmöglichkeiten hat sich spätestens seit der Arbeit von Greenberg (1960) auch in diesem Teilbereich der Sprachwissenschaft die Notwendigkeit einer quantitativen Untersuchung des Grades an Analytizität / Synthetizität durchgesetzt. Der konkrete Vorschlag von Greenberg (1960:187) lautet, die Anzahl von Morphemen pro Wort als „Syntheseindex“ (Morpheme / Wort) heranzuziehen. Darüber hinaus geht Greenberg (ibid.) noch einen Schritt weiter und schlägt eine – im Grunde allerdings völlig willkürliche – Einteilung vor, wonach eine Sprache mit einer durchschnittlichen Wortlänge (in der Anzahl von Morphemen) zwischen 1,00 und 1,99 als analytisch, eine Sprache mit einer Wortlänge zwischen 2,00 und 2,99 aber als synthetisch gilt, während eine Sprache ab einer Wortlänge von 3,00 als polysynthetisch anzusehen sei.

Es ist hier nicht der Ort, die konzeptuelle und methodologische Schwäche¹ dieser Art quantitativer Erfassung zu diskutieren, hervorzuheben ist jedoch, dass die Wortlänge in der Anzahl von Morphemen als entscheidender Parameter des Grades an Analytizität / Synthetizität angesehen wird. Eine ausführliche Übersicht über weitere mögliche Messungen des Synthesegrades findet sich in Altmann / Leheldt (1973:108ff.). Insgesamt dreht² sich aber der Versuch einer quantitativen Erfassung des Grades der Analytizität und Synthetizität hauptsächlich um die Wortlänge nach der Anzahl von Morphemen. Dabei sind allerdings drei zentrale Faktoren zu berücksichtigen bzw. zu differenzieren:

1. Das Problem der Definition von „Wort“ bzw. „Wortform“: D.h. im Vorfeld jeder Untersuchung ist zu klären, ob man von orthographischen, phonetischen, phonologischen oder morphologischen Kriterien bei der Wortbestimmung ausgeht.
2. Für die Untersuchung der Wortlänge ist von entscheidender Bedeutung, dass diese innerhalb einer Sprache in erster Linie von der Textsorte bzw. vom Funktionalstil abhängt. Dieser Zusammenhang konnte in systematischen Untersuchungen zum Slowenischen und Russischen bereits mehrfach nachgewiesen werden (vgl. Grzybek, Kelih 2005, Grzybek et al. 2005, Kelih et al. 2005), so dass bei einer vergleichenden Untersuchung mehrerer Sprachen insbesondere auf diesen Faktor zu achten ist. Bei einer Messung der Wortlänge auf syntagmatischer Ebene, d.h. auf Textebene, ist darüber hinaus zu beachten, dass zumindest zwei Ebenen zu unterscheiden sind: Die Messung eines jeden im Text vorkommenden Wortes (Wortformen-Tokens), oder aber die von Wortformen-Types, d.h. die Anzahl von Wortformen ohne Berücksichtigung der Frequenz im Text. Darüber hinaus eröffnen sich bei einer Lemmatisierung der Texte noch viele weitere Untersuchungsmöglichkeiten.
3. Die Wortlänge kann in verschiedenen Einheiten gemessen werden, wenn auch vor allem in Bezug auf die Messung der Analytizität / Synthetizität das Morphem als Maßeinheit vorgeschlagen wird. Darüber hinaus ist zu bedenken, dass eine Messung nach der Anzahl von Graphemen / Phonemen bzw. sogar Silben ebenfalls durchaus plausibel erscheint, da es letztlich um die Länge des sprachlichen Ausdrucks geht, die Informationen über den Grad an Analytizität / Synthetizität beinhaltet.

3. Parallel-Texte als adäquate Untersuchungsbasis

In Anbetracht der Tatsache, dass Textsorten und Funktionalstile einen großen Einfluss auf die Wortlänge in Texten haben, ist davon auszugehen, dass bei der Bestimmung der Wortlänge in vergleichenden Untersuchungen große Bedeutung der Textauswahl zugemessen werden sollte. Aus linguistischer Sicht ergibt sich für vergleichende Studien die Möglichkeit der Untersuchung von Paralleltexten, d.h. von Übersetzungen ei-

¹ Zur Anwendung des Syntheseindex vgl. Cowgill (1966). Spezifische Weiterentwicklungen einer derartigen Quantifizierung gehen auf Krupa (1965) und Krupa, Altmann (1966) zurück. Es wird vorgeschlagen den Syntheseindex in der Form Wörter / Morpheme zu berechnen und erstmals werden die Korrelationen zwischen weiteren Kenngrößen (wie in etwa des Ausmaßes der Agglutination, des Derivationsverhaltens usw.) hergeleitet und empirisch auf 20 genetisch nicht verwandte Sprachen angewandt. Darüber hinaus wird der von Greenberg (1960) vorgeschlagene Index auf das Intervall <0,1> normiert, so dass ein transparenter Vergleich über mehrere Sprachen hinweg möglich ist.

² Vorschläge zu einer Erweiterung (Kelemen 1970, Slavíčková 1968) beziehen sich auf weitere ausgewählte morphologische Eigenschaften, wobei ebenfalls die Häufigkeit von Morphemen (sei es in der Form von nicht weiter differenzierten Proportionen von analytischen und synthetischen Wortformen oder allgemein von analytischen Konstruktionen usw.) in Betracht gezogen wird.

nes bestimmten Textes. Auch wenn die Verwendung von Paralleltexten für sprachtypologische Vergleiche zuweilen aufgrund unterschiedlicher Argumente (Qualität der Übersetzung, fehlende Textteile, Problem der „translationese“, fehlende Äquivalente bzw. Explizitheit in Übersetzungen usw.) zuweilen auf Widerstand stößt (vgl. dazu Mauranen 2002, Wälchli 2007), ist darin dennoch ein plausibler Versuch zu sehen, eine möglichst umfassende Homogenisierung des verwendeten Materials für vergleichende Studien zu erreichen.

3.1 Untersuchungsschritte

Im Mittelpunkt der vorzunehmenden Analyse stehen das Russische, Mazedonische, Bulgarische, Slowakische und Slowenische. Die Berechnung der durchschnittlichen Wortlänge nach der Anzahl von Morphemen wird in zwei unterschiedlichen Korpora durchgeführt: Zum einen im russischen Roman „Kak zakaljalas' stal' (KZS)“ von N. Ostrovskij (vgl. dazu Kelih 2009) und zum anderen in „Master i Margarita (MiM)“ von M. Bulgakov und deren Übersetzung in die genannten slawischen Sprachen. Es wird dabei jeweils ausschließlich das erste Kapitel untersucht. Um nun zur Untersuchung selbst zu kommen:

1. Die genannten Texte werden einer Tagging-Prozedur unterzogen, indem Kapitelüberschriften, Abkürzungen, Zahlen usw. nach einem einheitlichen Prinzip bearbeitet werden.
2. Die Wortform wird nach orthographischen Kriterien (vgl. dazu detailliert Kelih 2007) bestimmt, d.h. jede Sequenz, die durch ein Leerzeichen getrennt ist, wird als eine Wortform angesehen. Der Bindestrich gilt dabei als wortabgrenzendes Mittel. Es werden allerdings ausschließlich Wortformen-Typen analysiert, d.h. die Wortformen ohne Berücksichtigung ihrer Frequenz im Text.
3. In den erstellten Listen von Wortformen-Typen wird manuell die Anzahl der Morpheme pro Wortform bestimmt. Bei der Segmentierung empfiehlt sich die Orientierung an verfügbaren morphologischen und einsprachigen Wörterbüchern der jeweiligen Sprache. Bei der Segmentierung werden Null-Morpheme nicht gezählt. Es werden allerdings nur diejenigen Wortformen ausgewertet, deren Frequenz größer ist als eins. Ein solches Vorgehen folgt daraus, dass anderenorts eine gesonderte Analyse der sogenannten hapax legomena erfolgt (Kelih 2011), die in einen erst kürzlich vorgeschlagenen Analytismus-Index Eingang finden (vgl. Popescu, Altmann 2008).

3.2 Wortlänge als Parameter

Bevor auf die Bedeutung der Wortlänge für die morphologische Klassifizierung eingegangen werden kann, ist auf die bisherige Zuordnung slawischer Sprachen zu dem einen oder anderen morphologischen Typus (analytisch / synthetisch) einzugehen. Dazu findet sich bei Hinrichs (2000a:91) folgende Einschätzung:

„In der Slavistik und in der Sprachtypologie geht man im allgemeinen davon aus, dass die slawischen Sprachen Bulgarisch und Mazedonisch ‚analytische‘ Sprachen seien, Russisch und Polnisch dagegen ausgesprochen ‚synthetische‘. Dies ist ein traditioneller Topos in der slavischen Sprachwissenschaft und fester Bestandteil des slavistischen Grundwissens. Zwischen beiden Polen liegt allenfalls das Serbokroatische als eine typologische Mischsprache – eine synthetische Sprache (Morphologie des Nomens) mit analytischen Zügen (Morphologie des Verbums).“

Der Vorteil einer quantitativen Analyse gegenüber einer Interpretation singulärer Eigenschaften von Sprachen ist vor allem darin zu sehen, dass sich eine generelle Aussage über die untersuchten Sprachen treffen lässt. Die Wortlänge kann dabei als zumindest

grober Gradmesser der Analytizität verstanden werden. Die Begründung dafür lautet folgendermaßen: Analytische Sprache sollten im Grunde genommen einen hohen Ausnutzungsgrad bei synsemantischen Wörtern zeigen, die als ein „autonomer“ Ausdruck kodiert werden und daher insgesamt aus grammatikalischen und syntaktischen Gründen häufiger vorkommen müssen. Daher verfügen analytische Sprache über eine geringere Wortlänge als synthetische Sprachen.

Darüber hinaus ist zu bedenken, dass in den hier untersuchten slawischen Sprachen auch der „morphologische Reichtum“ im Sinne der jeweiligen Ausprägung der Flexionsaktivität einen direkten Einfluss auf die durchschnittliche Wortlänge hat. So kann in Bezug auf das hier untersuchte Mazedonische und das Bulgarische festgehalten werden, dass der fast vollständige Kasusverlust zu einer absoluten Reduktion von unterschiedlichen Wortformen-Typen führt. Die Verringerung von Kasusformen wird aber durch einen erhöhten Einsatz von Präpositionalkonstruktionen kompensiert, was wiederum einen Einfluss auf die Häufigkeit von synsemantischen Wörtern zeigt. Gleichzeitig verfügen das Bulgarische und das Mazedonische über einen synthetisch ausgedrückten postpositiven Artikel, welcher wiederum zu einer Längung von einzelnen Wortformen führt.

Insgesamt ist daher die Wortlänge ein brauchbarer Gradmesser, da so die gesamte morphologische Struktur eines Textes erfasst und gleichzeitig ein Einblick in das Verhältnis zwischen der Anzahl an Wortformen-Typen und Anzahl an Morphemen gegeben wird.

3.3 Wortlänge in Paralleltexten: KZS

Zu beginnen ist mit den Resultaten einer Analyse der durchschnittlichen Wortlänge im ersten Kapitel des KZS-Korpus. In Tabelle 1 finden sich die entsprechenden Rohdaten. Wie aus der verwendeten Datenbasis ersichtlich ist, handelt es sich um ein relativ kleines Korpus, welches lediglich einen ersten Einblick in die morphologische Struktur der untersuchten Sprachen geben soll.

Tabelle 1: Rohdaten zum KZS- Korpus (1. Kapitel)

Sprache	Types	Morpheme	Morphemlänge
Slowenisch	574	1261	2,1969
Mazedonisch	517	1282	2,4797
Slowakisch	486	1072	2,2058
Russisch	480	1151	2,3979
Bulgarisch	520	1246	2,3962

Ein erster empirischer Befund aus den angeführten Daten ist, dass sich die Anzahl von Types in den fünf von uns untersuchten Sprachen unterscheidet. Das Slowenische verfügt über die höchste Anzahl (574), gefolgt von Mazedonisch und Bulgarisch, die beide in etwa die gleiche Anzahl von Types haben ($\approx 518,5$). Dies gilt auch für das Slowakische und das Russische (≈ 483).

Um damit zur Morphemlänge in den untersuchten slawischen Sprachen zu kommen: Wie der Tabelle 1 zu entnehmen ist, hat das Slowenische mit durchschnittlich 2,20 Morphemen pro Type die geringste Wortlänge, knapp gefolgt vom Slowakischen mit 2,21. Das Russische, gemeinhin als Sprache mit synthetischen Zügen bezeichnet, scheint hier mit 2,39 Morphemen pro Wortform eine mittlere Position einzunehmen. Diese Position teilt es sich mit dem Bulgarischen, welches ebenfalls eine mittlere Wortlänge (in der Anzahl von Morphemen) von 2,39 hat. Das Mazedonische hat im Durchschnitt 2,48 Morpheme und damit die höchste Wortlänge innerhalb der hier untersuchten slawischen

Sprachen. So interessant auch die deskriptiven Befunde sind, sie erübrigen entsprechende statistische Signifikanztests nicht, weil nur diese Auskunft geben über die Signifikanz der Unterschiede zwischen den einzelnen Sprachen.

Da die eruierten Wortlängenhäufigkeiten nicht der Normalverteilung³ folgen, wird ein nichtparametrischer Test (Kruskal-Wallis-Test) verwendet, um zu untersuchen, ob sich die Sprachen hinsichtlich ihrer zentralen Tendenz (des Medians) signifikant voneinander unterscheiden oder nicht. Dabei zeigt sich folgendes Ergebnis (vgl. Tabelle 2): Keine signifikanten Unterschiede ($p > 0,05$) bestehen zwischen Slowenisch und Slowakisch. Zwischen dem Slowenischen und allen anderen slawischen Sprachen (Russisch, Bulgarisch, Mazedonisch) ist allerdings ein signifikanter Unterschied festzustellen. Des Weiteren gibt es zwischen Slowakisch und Bulgarisch einen Unterschied von ($p < 0,05$). Das Russische unterscheidet sich weder vom Bulgarischen noch vom Mazedonischen ($p > 0,05$) und dies gilt auch für das Bulgarische und das Mazedonische.

Tabelle 2: Ergebnisse des Signifikanztests für den KZS-Korpus (Kruskal-Wallis-Test)

Sprachenpaare		p	Sprachenpaare		p
Slowenisch	Slowakisch	0,415782	Slowakisch	Bulgarisch	0,000187*
	Russisch	0,006315*	Russisch	Bulgarisch	0,744495
	Bulgarisch	0,001162*	Russisch	Mazedonisch	0,333693
	Mazedonisch	0,000012*	Bulgarisch	Mazedonisch	0,250825

Mit anderen Worten: Slowenisch und Slowakisch zeigen gegenüber Bulgarisch, Mazedonisch und Russisch hinsichtlich der durchschnittlichen Anzahl der Morpheme pro Wort ein jeweils gruppenspezifisches Verhalten. Vgl. dazu auch Abb. 1.

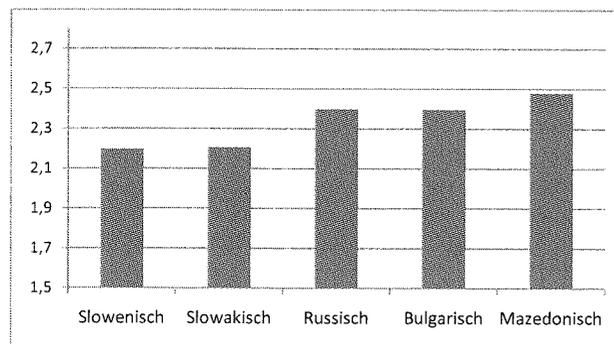


Abb. 1: Durchschnittliche Wortlänge in Morphemen im KZS-Korpus (Wortformen-Types)

Vor einer abschließenden Verallgemeinerung sind die hier erzielten Ergebnisse durch ein weiteres Kontroll-Korpus abzusichern. Hierzu werden der russische Roman „Master i Margarita“ (MiM) und die jeweiligen Übersetzungen ins Bulgarische, Mazedonische, Slowenische und Slowakische nach demselben Schema wie das KZS-Korpus untersucht.

³ Das wurde durch den für diese Stichprobengröße geeigneten Shapiro-Wilk- und den Kolmogorov-Smirnov-Test überprüft. Zu Details und der Anwendung des Verfahrens in der Wortlängenanalyse vgl. Antić, Kelih, Grzybek (2006).

3.4 Wortlänge in Paralleltextrn: MiM

Die Untersuchung der Wortlänge (nach der Anzahl von Morphemen) zeigt in dem zweiten untersuchten Korpus folgende Ergebnisse (vgl. Tabelle 3). Das Slowenische verfügt wiederum über die geringste Wortlänge nach der Anzahl von Morphemen (2,27), knapp gefolgt vom Slowakischen (2,32). Im MiM-Korpus hat das Russische eine geringfügig höhere Wortlänge (2,56) als das Bulgarische (2,50) und das Mazedonische verfügt auch in dieser Analyse über die höchste Wortlänge von durchschnittlich 2,62 Morphemen pro Wortformen-Types. (Vgl. dazu auch Abb. 2).

Tabelle 3: Auswertungen von „Master i Margarita“ (1. Kapitel)

Sprache	Types	Morpheme	Morphemlänge
Slowenisch	436	991	2,2729
Mazedonisch	449	1175	2,6169
Slowakisch	406	940	2,3153
Russisch	402	1031	2,5647
Bulgarisch	446	1117	2,5045

Wiederum sind vor einer abschließenden Interpretation der Ergebnisse entsprechende Signifikanztests durchzuführen. Wie der Tabelle 4 zu entnehmen ist, ergeben sich im Grunde genommen die gleichen Resultate wie beim KZS-Korpus. Das Slowenische und das Slowakische unterscheiden sich nicht signifikant, während das Bulgarische, Mazedonische und Russische sich deutlich vom Slowenischen unterscheiden. Auch unterscheidet sich das Slowakische signifikant vom Bulgarischen. Es ergeben sich allerdings keine großen Unterschiede zwischen Russisch und Bulgarisch, Russisch und Mazedonisch sowie Bulgarisch und Mazedonisch. Mit anderen Worten: Es ergibt sich sowohl im MiM-Korpus als auch im KZS-Korpus das gleiche Gesamtbild.

Tabelle 4: Ergebnisse des Signifikanztestes für das KZS-Korpus (Kruskal-Wallis-Test)

Sprachenpaare		p	Sprachenpaare		p
Slowenisch	Slowakisch	0,488330	Slowakisch	Bulgarisch	0,000442*
	Russisch	0,004624*	Russisch	Bulgarisch	0,336709
	Bulgarisch	0,000044*	Russisch	Mazedonisch	0,349412
	Mazedonisch	0,000047*	Bulgarisch	Mazedonisch	0,986954

Damit ist als einstweiliges Zwischenergebnis festzuhalten, dass wiederum Slowenisch / Slowakisch als analytische Sprachen in Erscheinung treten, während Bulgarisch, Mazedonisch und Russisch im Gegensatz dazu als synthetisch zu bezeichnen sind.

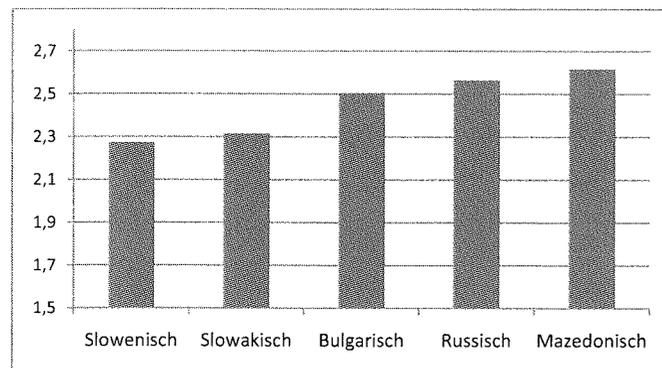


Abb. 2: Durchschnittliche Wortlänge in Morphemen im MiM-Korpus (Wortformen-Types)

4. Abschließende Interpretation

Die vorliegende Untersuchung der durchschnittlichen Wortlänge nach der Anzahl von Morphemen in slawischen Paralleltexten lässt eine Vielzahl von vorläufigen Verallgemeinerungen zu. Interpretiert man diesen Syntheseindex im Sinne von Greenberg (1960), so sind alle hier untersuchten slawischen Sprachen als synthetisch einzuschätzen, da in allen Sprachen eine durchschnittliche Wortlänge > 2,00 zu beobachten ist. Innerhalb dieses Spektrums lassen sich allerdings zwei Gruppen unterscheiden: Eine etwas weniger synthetische (= analytische) Gruppe mit dem Slowenischen und dem Slowakischen und eine synthetische Gruppe mit den Sprachen Bulgarisch, Mazedonisch und Russisch. Der Befund, dass das Bulgarische, das Mazedonische und das Russische aufgrund dieses Parameters gleichsam zu einem morphologischen Typus gehören, mag aus slawistischer Sicht auf den ersten Blick überraschen. Die eingangs zitierte Lehrmeinung, wonach Russisch innerhalb der slawischen Sprachen als synthetisch und Bulgarisch / Mazedonisch als analytisch gelten, kann sich im Grunde nur auf die paradigmatische Ebene und hier im Speziellen auf die Ausprägung der Kasusmorphologie beziehen. Eine syntagmatische Analyse, also eine Textanalyse, zeigt dem gegenüber ein anderes Bild. Offenbar spielt für diesen Befund, d.h. für die relativ hohe Wortlänge des Bulgarischen und Mazedonischen im Gegensatz zu anderen Sprachen, hauptsächlich der nachgestellte bestimmte Artikel eine Rolle, weil der eine Verlängerung von Wortformen nach sich zieht. In Anbetracht der eindeutigen Ergebnisse wird man nicht umhin kommen, das Bulgarische und das Mazedonische auf der Basis einer syntagmatischen Wortformen-Analyse gemeinsam mit dem Russischen als „synthetische Sprachen“ zu bezeichnen. Das Slowenische und das Slowakische erweisen sich bei dieser Form der Untersuchung als analytische Sprachen, zumindest unterscheidet sich die durchschnittliche Wortlänge nach der Anzahl von Morphemen signifikant vom Russischen, Bulgarischen und Mazedonischen. Unterstrichen werden muss jedoch noch einmal, dass diese Ergebnisse vermutlich nur für eine Analyse von Wortformen-Types gelten. Bei der Analyse von Wortformen-Tokens bzw. einer zusätzlichen Analyse von im Text vorkommenden hapax legomena wird sich sicherlich ein anderes Gesamtbild ergeben. Dieses sollte aber nicht darüber hinwegtäuschen, dass 1. eine quantitative Analyse von Analytismus / Synthetismus einen tieferen Einblick in die morphologische Struktur slawischer Sprachen erlaubt und sich 2. Parallel-Texte als eine brauchbare Untersuchungsbasis erweisen.

Literatur

- Altmann, G., Lehfeldt, W. 1973. *Allgemeine Sprachtypologie*. München.
- Antić, G., Kelih, E., Grzybek, P. 2006. Zero-syllable Words in Determining Word Length. In: Grzybek, P. (Hrsg.): *Contributions to the Science of Text and Language. Word Length Studies and Related Issues*. Dordrecht, 117-156.
- Bossong, G. 2004. Analytizität und Synthesizität. Kasus und Adpositionen im typologischen Vergleich. In: Hinrichs, U. (Hrsg.): *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*. Wiesbaden, 431-452.
- Coseriu, E. 1972. Über die Sprachtypologie Wilhelm von Humboldts. Ein Beitrag zur Kritik der sprachwissenschaftlichen Überlieferung. In: Höhle, J. (Hrsg.): *Beiträge zur vergleichenden Literaturgeschichte. Festschrift für Kurt Wais zum 65. Geburtstag*. Tübingen, 107-135.
- Cowgill, W. 1966. A search for universals in indo-european diachronic morphology. In: Greenberg, J. H. (Hrsg.): *Universals of language*. 2nd edition. Cambridge-MA, 114-141.
- Dixon, R. M. W.; Aikhenvald, A. Y. (Hrsg.) 2002a. *Word. A cross linguistic typology*. Cambridge u.a.
- Dixon, R. M. W.; Aikhenvald, A. Y. 2002b. Word: a typological framework. In: Dixon, R. M. W.; Aikhenvald, A. Y. (Hrsg.): *Word. A cross linguistic typology*. Cambridge u.a., 1-41.
- Grzybek, P., Kelih, E. 2005. Empirische Textsemiotik und quantitative Text-Typologie. In: Bernard, J., Fikfak, J., Grzybek, P. (Hrsg.): *Text & Reality. Text & Wirklichkeit*. Ljubljana-Wien-Graz: ZRC, 95-120.
- Grzybek, P., Stadlober, E., Kelih, E., Antić, G. 2005. Quantitative Text Typology: The Impact of Word Length. In: Weihs, C., Gaul, W. (Hrsg.): *Classification. The Ubiquitous Challenge*. Heidelberg, 53-64.
- Hinrichs, U. 2000. Prolegomena zu einer Theorie des Analytismus I. Anhand der Sprachen in Ost- und Südosteuropa. In: Hinrichs, U., Büttner, U. (Hrsg.): *Die Südosteuropa-Wissenschaften im neuen Jahrhundert: Akten der Tagung vom 16.-19.10.1999 an der Universität Leipzig*. Wiesbaden, 83-105.
- Hinrichs, U. 2004. Vorbemerkungen zum analytischen Sprachtyp in Europa. In: Hinrichs, U. (Hrsg.): *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*. Wiesbaden, 17-31.
- Kelemen, J. 1970. Sprachtypologie und Sprachstatistik. In: Dezső, L., Hajdú, P. (Hrsg.): *Theoretical problems of typology and the northern Eurasian languages*. Amsterdam, 53-63.
- Kelih, E. 2009. Slawisches Parallel-Textkorpus: Projektvorstellung von "Kak zakaljalas' stal' (KZS)". In: Kelih, E., Levickij, V.V., Altmann, G. (Hrsg.): *Methods of Text Analysis. Metody analizu tekstu*. Černivci, 106-124.
- Kelih, E. 2007. Zur Frage der Wortdefinitionen in Wortlängenuntersuchungen. In: Kaliuščenko, V., Köhler, R., Levickij V. (Hrsg.): *Problems of Typological and Quantitative Lexicology*. Chernivtsi, 91-105.
- Kelih, E., Antić, G., Grzybek, P., Stadlober, E. 2005. Classification of Author and/or Genre? The Impact of Word Length. In: Weihs, C., Gaul, W. (Hrsg.): *Classification. The Ubiquitous Challenge*. Heidelberg-New York, 498-505.
- Krámský, J. 1969. *The word as a linguistic unit*. The Hague.
- Krupa, V. 1965. On quantification of typology. *Linguistics* 3, 12, 31-36.
- Krupa, V., Altmann, G. 1966. Relations between typological indices. *Linguistics* 4, 24, 29-37.
- Popescu, I., Altmann, G. 2008. Hapax legomena and Language Typology. *Journal of Quantitative Linguistics*, Jg. 15, H. 4, 370-378.
- Reiter, N. 2004. Die semantische Dimension des sogenannten Analytismus. In: Hinrichs, U. (Hrsg.): *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*. Wiesbaden, 87-102.
- Schwegler, A. 1990. *Analyticity and diachronic perspective with special reference to romance languages*. Berlin-New York.
- Slavičková, E. 1968. Towards a typological evaluation of related languages. *Travaux Linguistiques de Prague*, H. 3, 281-289.
- Wälchli, B. 2007. Advantages and disadvantages of using parallel texts in typological investigations. *Sprachtypologie und Universalienforschung* 60, 2, 118-134.

Резюме

В данной статье исследуется степень аналитизма/синтетизма в болгарском, македонском, русском, словацком и словенском языках. Степень аналитизма/синтетизма определяется длиной слова (количеством морфем). На основе параллельных текстов («Как закалялась сталь», «Мастер и Маргарита») показано, что словенский и словацкий являются аналитическими языками, а русский, болгарский и македонский – синтетическими.

Graz

(emmerich.kelih@uni-graz.at)

Emmerich Kelih

**DIE WELT DER SLAVEN
SAMMELBÄNDE · СБОРНИКИ**

Herausgegeben von
Peter Rehder (München) und Igor Smirnov (Konstanz)

Band 43



VERLAG OTTO SAGNER
München – Berlin 2011

**BEITRÄGE DER
EUROPÄISCHEN SLAVISTISCHEN
LINGUISTIK
(POLYSLAV)**

Band 14

Herausgegeben von
Katrin Bente Karl
Gertje Krumbholz
Marija Lazar



VERLAG OTTO SAGNER
München – Berlin 2011