



Analytismus/Synthetismus in slawischen Sprachen

Empirische Ergebnisse aus Paralleltextkorpora

Institut für Slawistik

emmerich.kelih@uni-graz.at

<http://www.uni-graz.at/emmerich.kelih/>

Überblick

1. Analytisch (A) vs. Synthetisch (S): Definition
2. State of the art in der slawischen Sprachwissenschaft
3. Fallbeispiele aus slawischen Sprachen
 - 3.1. Komparation im Slowenischen
 - 3.2. Futur im Serbischen/Kroatischen
 - 3.3. Präpositionalkonstruktionen im Russischen
3. (A) (S) in Paralleltextkorpora (Originaltexte und Übersetzungen)
4. Syntheseindex: Anzahl von Morphemen pro Wort
5. Pilotstudie: Empirische Ergebnisse
6. Zusammenfassung und Perspektiven

Morphologische Klassifikation:

isolierender, agglutinierender, inkorporierender und flektierender Typus ...



analytische vs. synthetische Verfahren

Analytisch:

grammatische Information wird außerhalb eines Voll-Lexems, d.h. als „autonome“ Wortform ausgedrückt. (Hinrichs 2000a)

oder:

Umverteilung von lexikalischer bzw. grammatischer Information von der morphologischen auf die syntaktische Ebene.

Beispiele: Futur I

‘wir werden sehen‘

Russisch: Synthetisch

po-smotr-im

PRÄF.ASP.FUT.-schauen-1PL

Serbisch: Analytisch

ħe-mo

će-mo

AUX.FUT-1PL.PRÄS

vide-mu

vide-ti

sehen-Inf.

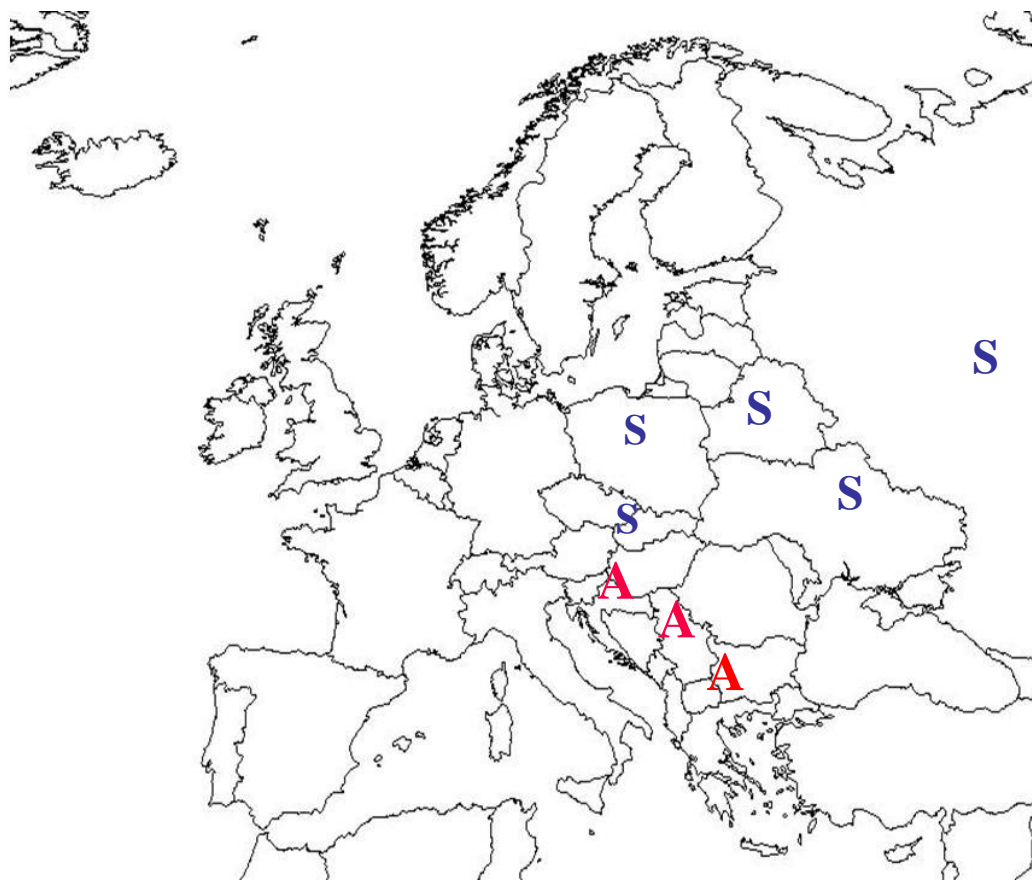
Arealtypologie: (Gvozdanović 2009)

A. das südliche Areal: **Analytisch**

- südöstliche Areal (Bulgarisch und Mazedonisch)
- Serbisch als Übergangsgebiet
- südwestliche Areal (Slowenisch, Kroatisch)

B. der nördliche Bereich: **Synthetisch**

- Westslawische
- Ostslawische



traditioneller Topos der slavischen Sprachwissenschaft (Hinrichs 2000)

Synthetisch

Russisch (Ostslawisch)
Polnisch (Westslawisch)

Serbokroatisch (B/K/S)
(Übergangsgebiet)

Analytisch
Bulgarisch
Mazedonisch
(Südostslawisch)

Offene Fragen und Probleme:

- Bulgarisch/Mazedonisch als analytische Sprachen?
- Serbisch als Übergangsgebiet?
- westslawische Sprachen insgesamt als synthetische Sprachen?

- welche sprachliche Ebene wird untersucht? (Grammatik vs. Text)
- welche Merkmale liegen dieser Zuordnung zugrunde?
- Verwendung von (A) und (S) innerhalb einer Sprache?

Fallbeispiel 1: Komparation im Slowenischen:

I. Synthetisch (bei häufig gebrauchten Adjektiven)

<i>lep-Ø</i> schön-M.NOM 'schön'	<i>lep-š-i</i> schön-COMP-M.DEF 'schöner'	<i>naj-lep-š-i</i> SUPERLAT-schön- COMP-M.DEF 'am schönsten'
<i>bogat-Ø</i> reich-M.NOM 'reich'	<i>bogat-ejš-i</i> reich-COMP-M.DEF. 'reicher'	<i>naj-bogat-ejš-i</i> SUPERLAT-reich- COMP-M.DEF. 'am reichsten'

II. Analytisch (i.d.R. bei mehrsilbigen Wörtern, Farbwörtern, Endung auf -en, -av, -ast, -a)

<i>zaželen-Ø</i> gewünscht-M.Nom 'gewünscht'	<i>bolj zaželen-Ø</i> COMP gewünscht-M.NOM. 'mehr gewünscht/gewünschter'	<i>najbolj zaželen-Ø</i> SUPERLAT -schön-M.NOM. 'am meisten gewünscht/gewünschttesten'
<i>muhast-Ø</i> lästig-M.Nom 'lästig'	<i>bolj muhast-Ø</i> COMP lästig-M.NOM 'lästiger'	<i>najbolj muhast-Ø</i> SUPERLAT-lästig-M.NOM. 'am lästigsten'

Fallbeispiel 2: Futur im Serbischen

I. Analytisch: Enklitische Form („Kurzform“) von хтети/hteti („wollen“) + Infinitiv

<i>mi</i>	<i>ће-мо</i>	<i>виде-ми</i>
<i>mi</i>	<i>ће-мо</i>	<i>виде-ти</i>
wir	AUX.FUT-1PL.PRÄS	sehen-Inf.
'wir werden sehen'		

II. Synthetisch: Infinitiv – ti + enkl. Form von хтети/hteti („wollen“)

<i>виде-ће-мо</i>	<i>раду-ће-ме</i>
<i>виде-ће-мо</i>	<i>ради-ће-те</i>
sehen.INF-AUX-1PL	arbeiten.INF.-AUX-2PL
'wir werden sehen'	'ihr werdet arbeiten'

aber: Infinitive auf –ći wie z.B.: доћи ћу/doći ću

III. Analytisch: Enklitische Form хтети/hteti („wollen“) + da-Konstruktion + Verb

<i>мо</i>	<i>ћемо</i>	<i>да</i>	<i>видимо</i>
to	će-mo	da	vid-imo
das/dies	werden-1Pl	da-Konjunktion	sehen-1Pl
'das werden wir sehen'			

Fallbeispiel 2a: Futur im Kroatischen

I. Analytisch: Infinitiv –i + enkl. Form von hteti

<i>videt</i>	<i>će-mo</i>	<i>radit</i>	<i>će-te</i>
sehen.INF	AUX-1PL	arbeiten.INF	AUX-2Pl
'wir werden sehen'		'ihr werdet arbeiten'	

aber: -ti bleibt erhalten, wenn nicht unmittelbar vor Hilfsverb

Resultate & Desiderata

- unterschiedliche Bildung im Serbischen und Kroatischen
- morphosyntaktische und phonologische Determiniertheit der Verwendung
- da-Konstruktion bei Futurbildung als „typisch“ Serbisch?
- keine systematischen Korpus-Untersuchungen
- Rolle von orthographischen Konventionen?
- Rolle von normativen Empfehlungen?

Fallbeispiel (3): Präpositionalkonstruktionen im Russischen

I. Analytisch

učebnik-∅

Lehrbuch-NOM.SG.
'Mathematiklehrbuch'

po

über.PRÄP.

matematik-e

Mathematik-DAT.SG.

plan-∅

Plan-Nom.Sg.
'Produktionsplan'

po

über.PRÄP.

vypusk-u

Produktion-DAT.SG.

II. Synthetisch

učebnik-∅

Lehrbuch-NOM.SG.
'Mathematiklehrbuch'

matematik-i

Mathematik-GEN.SG.

plan-∅

Plan-Nom.Sg.
'Produktionsplan'

vypusk-a

Produktion-Gen.Sg.

Zwischenresümee:

Innerhalb einer Sprache zum Ausdruck grammatischer Kategorien:

1. sowohl analytische als auch synthetische Verfahren
2. dialektale, stilistische, textsortenspezifische Unterschiede
3. phonologische bzw. morphosyntaktische Determiniertheit der Verwendung

weiterführende Fragen:

1. Welche Rolle bei der Bestimmung von (A) und (S) spielt eine Textanalyse?

2. Wie lässt sich der Grad von (A) und (S) ermitteln?

Neuansatz zu (A) (S) in slawischen Sprachen

1. Differenzierung der sprachlichen Bezugsebene

(a) paradigmatische Ebene: Vorhandensein von Kodierungsverfahren im Sprachsystem

(b) syntagmatische Ebene: **Text-Analyse: Paralleltexte**

„Kak zakaljalas‘ stal’/Wie der Stahl gehärtet wurde” (KZS)

„Master i Margarita/Meister und Margarita” (MiM) (Kelih 2009a, 2009b)

Selektive Sprachenauswahl:

- Russisch (Ostslawisch)
- Bulgarisch (Südostslawisch)
- Mazedonisch (Südostslawisch)
- Serbisch (“Übergangsgebiet”)
- Slowenisch (Südwestslawisch)
- Slowakisch (Westslawisch)

Wie lässt sich der Grad von (A) und (S) ermitteln?

- (a) Texte als Konglomerat unterschiedlicher Kodierungsverfahren (Komparation, Temporalsystem ...)
- (b) quantitative Charakterisierung: „Syntheseindex“ (Greenberg 1960)

Wortlänge (WOL): Anzahl von Morphemen (M) pro Wortform (W)

- **Je kürzer die Wortformen, desto analytischer eine Sprache.**
- **Je länger die Wortformen, desto synthetischer eine Sprache.**

Analytisch

učeb-nik-∅

Lehrbuch-NOM.SG.

'Mathematiklehrbuch'

po

über.PRÄP.

matematik-e

Mathematik-DAT.SG.

3 Wortformen

6 Morpheme

$M/W = 6/3 = \mathbf{2.00}$

Synthetisch

učeb-nik-∅

Lehrbuch-NOM.SG.

'Mathematiklehrbuch'

matematik-i

Mathematik-GEN.SG.

2 Wortformen

5 Morpheme

$M/W = 5/2 = \mathbf{2.5}$

Pilotstudie: Wortlänge (Anzahl von Morphemen) in Paralleltexten

jeweils 1. Kapitel:

- I. „Kak zakaljalas‘ stal’/Wie der Stahl gehärtet wurde (KZS)
- II. „Master i Margarita/Meister und Margarita” (MiM)

- orthographische Wortform (Leerstellenkriterium)
- „Wortlänge“ (WOL): Anzahl von Morphemen

učeb-nik-Ø

Lehrbuch-NOM.SG.
'Mathematiklehrbuch'

po

über.PRÄP.

matematik-e

Mathematik-DAT.SG.

3 Wortformen

6 Morpheme

$M/W = 6/3 = 2.00$

Fallbeispiel: Phrase aus KZS

‘gleich werden wir sehen‘

Russisch

мы *сей-час* *по-смотрим-им.*
my *sej-čas* *po-smotr-im.*
 wir 1PL gleich.NOM.SG.M.-Zeit.NOM.SG.M. nach.PERF.-schauen-1PL

Serbisch

сад *ће-мо* *виде-му.*
sad *će-mo* *vide-ti*
 gleich-ADV. werden-1PL. sehen-INF.

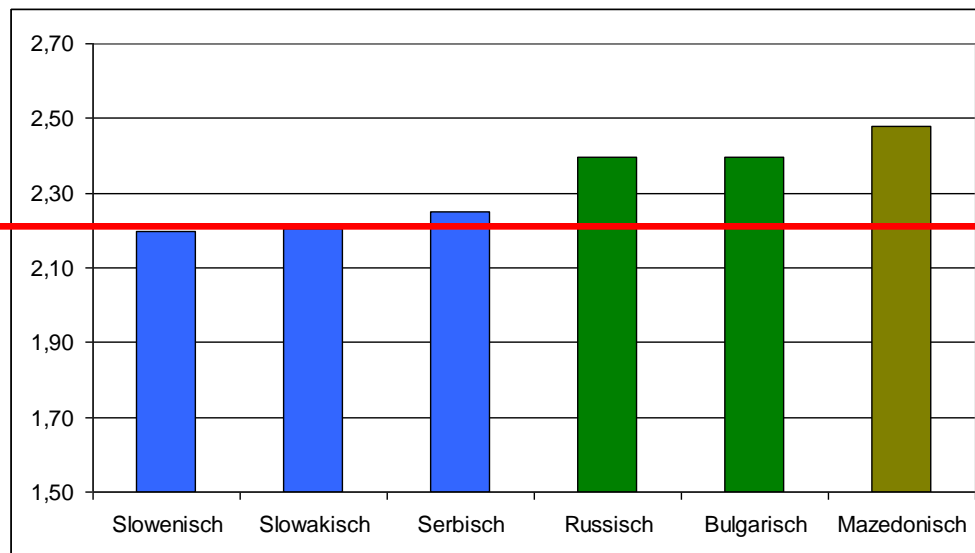
Bulgarisch

сега *ще* *вид-им.*
sega *šte* *vid-im.*
 gleich-ADV. AUX sehen-1PL

Sprache	Anzahl Wortformen	Morpheme	M/W
Russisch	3	6	2.00
Serbisch	3	5	1.66
Bulgarisch	3	4	1.33

“KZS”’: Wortlänge (Types) 1. Kapitel

Sprache	Types	Morpheme	Wortlänge
Slowenisch	574	1261	2,20
Slowakisch	486	1072	2,21
Serbisch	537	1209	2,25
Russisch	480	1151	2,40
Bulgarisch	520	1246	2,40
Mazedonisch	517	1282	2,48



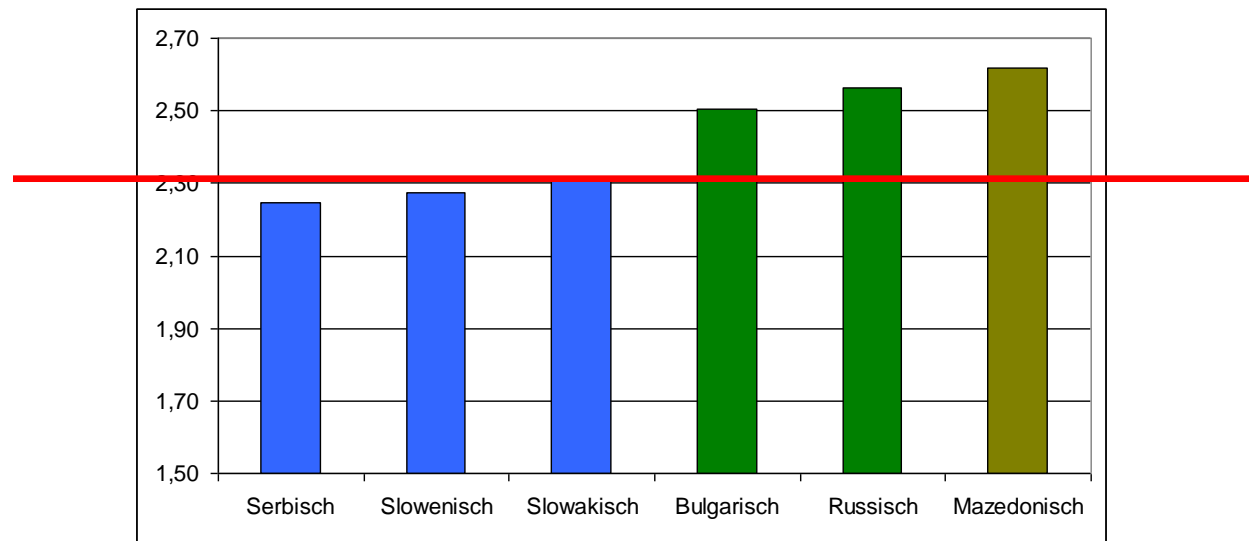
→ Je länger das Wort, desto synthetischer die Sprache.

→ signifikante Unterschiede (nichtparametrische Test)

[Slowenisch, Slowakisch, Serbisch] vs. [Mazedonisch, Russisch, Bulgarisch]

“MiM” : Wortlänge (Types) 1. Kapitel

Sprache	Types	Morpheme	Wortlänge
Serbisch	444	997	2,25
Slowenisch	436	991	2,27
Slowakisch	406	940	2,32
Bulgarisch	446	1117	2,50
Russisch	402	1031	2,56
Mazedonisch	449	1175	2,62



→signifikante Unterschiede:

[Slowenisch, Slowakisch, Serbisch] vs. [Mazedonisch, Russisch, Bulgarisch]

traditioneller Topos in der slawischen Sprachwissenschaft

Synthetisch
Russisch (Ostslawisch)
Polnisch (Westslawisch)

Serbokroatisch (B/K/S)
(Übergangsgebiet)

Analytisch
Bulgarisch
Mazedonisch
(Südostslawisch)

synthetisch

Syntheseindex: (A) und (S)

analytisch



Mazedonisch, Bulgarisch, Russisch

Slowakisch, Serbisch, Slowenisch

Vorläufige Ergebnisse:

- Mazedonisch/Bulgarisch können kaum als analytisch bezeichnet werden

Erklärung: Längung von Wortformen durch postpositiven Artikel

Slowenisch, Slowakisch als analytische Sprachen

- Serbisch als analytische Sprache
- Russisch als synthetische Sprache

Resultate

1. Aussagen über (A) (S) sind nach sprachlichen Untersuchungsebenen zu differenzieren
2. Parallel-Texte als geeignete Basis für die Untersuchung von (A) (S)
3. Wortlänge in Morphemen als adäquater Indikator
4. neue arealtypologische Gliederung:
[Mazedonisch, Bulgarisch, Russisch] vs. [Slowenisch, Serbisch, Slowakisch]

Perspektiven

1. Analyse weiterer Paralleltexte
2. Anwendung von unterschiedlichen Wortdefinitionen
3. Analyse des Basiswortschatz (Wortlisten)
4. weitere Indikatoren (morphologische Reichtum, „Hapax Legomena“)
5. Wechselbeziehungen zu anderen sprachlichen Ebenen

Literatur (Auswahl)

Coseriu, Eugenio (1972): Über die Sprachtypologie Wilhelm von Humboldts. Ein Beitrag zur Kritik der sprachwissenschaftlichen Überlieferung. In: Hösle, Johannes (Hg.): Beiträge zur vergleichenden Literaturgeschichte. Festschrift für Kurt Wais zum 65. Geburtstag. Tübingen: Niemeyer, S. 107–135.

Greenberg, Joseph H. (1960): A quantitative approach to the morphological typology of language. In: International Journal of American Linguistics, H. 26, S. 178–194.

Gvozdanović, Jadranka (2009): Synthetismus und Analytismus im Slavischen. In: Kempgen, Sebastian; Kosta, Peter; Berger, Tilman Gutschmidt Karl (Hg.): Die slavischen Sprachen. Teilband 1. Berlin/New York: Mouton De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 32), S. 129–142.

Hinrichs, Uwe (2000a): Prolegomena zu einer Theorie des Analytismus I. Anhand der Sprachen in Ost- und Südosteuropa. In: Hinrichs, Uwe; Büttner, Uwe (Hg.), S. 83–105.

Hinrichs, Uwe (2000b): Prolegomena zu einer Theorie des Analytismus II anhand der Sprachen in Ost- und Südosteuropa. In: Hinrichs, Uwe; Büttner, Uwe (Hg.), S. 107–128.

Hinrichs, Uwe; Büttner, Uwe (Hg.): Die Südosteuropa-Wissenschaften im neuen Jahrhundert: Akten der Tagung vom 16. - 19.10.1999 an der Universität Leipzig. Wiesbaden: Harrassowitz, S. 107–128.

Hinrichs, Uwe (2004): Zum Analytismus im Serbischen. In: Hinrichs, Uwe (Hg.): Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp. Wiesbaden: Harrassowitz (Eurolinguistische Arbeiten, 1), S. 293–302.

Kelih, Emmerich (2009a): Slawisches Parallel-Textkorpus: Projektvorstellung von "Kak zakaljalas' stal' (KZS)". In: Kelih, Emmerich; Levickij, V.V.; Altmann, Gabriel (Hg.): Methods of Text Analysis. Metody analizu tekstu. Černivci: ČNU, S. 106–124.

Kelih, Emmerich (2009): Preliminary analysis of a Slavic parallel corpus. In: Levická, Jana; Garabík, Radovan (Hg.): NLP, Corpus Linguistics, Corpus Based Grammar Research. Fifth International Conference Smolenice, Slovakia, 25-27 November 2009. Proceedings. Bratislava: Tribun, S. 175–183.

Hvala za vašo pozornost!

Спасибо за ваше внимание!

Vielen Dank für Ihre Aufmerksamkeit !

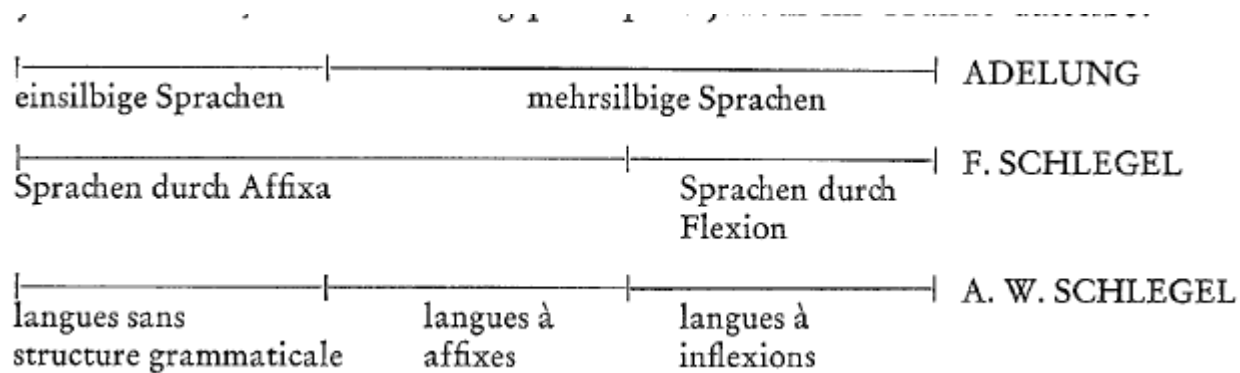
Postponitiver Artikel im Bulgarischen

ЧУЖДЕНЕЦЪТ	čuždenečăt	der Fremde
ПОЕТЪТ	poetăt	der Poet
РЕДАКТОРЪТ	redaktorăt	der Redakteur

Postponitiver Artikel im Mazedonischen

СТРАНЕЦОТ	stranecot	der Fremde
ПОЕТОТ	poetot	der Poet
УРЕДНИКОТ	urednikot	der Redakteur

Ein historischer Exkurs: Morphologische Klassifikation



Coseriu (1972)

Sukzessive Spezifizierung (u.a. Wilhelm von Humboldt)

- morphologisch „formlosen“ vs. morphologisch „geformte“
- isolierender, agglutinierender, inkorporierender und flektierender Typus ...



analytische vs. synthetische Verfahren

Probleme:

1. Wort und Wortdefinition
2. Frage von Wortgrenzen
3. kognitive Bedeutung des Wortbegriffes
4. Umgang in cross-linguistischen Untersuchungen
 - radikale Konsequenz: aufgrund fehlender universal gültigen Bestimmung des Begriffes Wort bzw. Wortgrenze ist, der völlige Verzicht auf eine Klassifikation auf analytische und synthetische Sprachen (Schwegler 1990: 46)
 - absolute Gültigkeit von Wortgrenzen ist ein theoretisches Konstrukt der Sprachtypologie (Haarman 2004: 73)
 - „Much that has been written about the word is decidedly eurocentric“.
(Dixon/Aikhenvald 2002: 2)

Einige Probleme mit Untersuchung von Parallel-Texten

- fehlende/nicht übersetzte Textteile (alignment)
- Vereinfachungen (auf allen sprachlichen Ebenen, insbesondere syntaktische und lexikalische V.)
- Qualität der Übersetzungen („translationese“)
- Übersetzungsstrategie: Explizitheit erhöhen („explicitation“)

Paralleltextkorpora

I. „Kak zakaljalas‘ stal’ – Wie der Stahl gehärtet wurde (KZS) in 12 slawischen Sprachen

- Roman aus den 30er Jahren (real. Sozialismus)
- Autor: N.A. Ostrovskij
- 10 Kapitel in 12 Sprachen:

Weißrussisch, Russisch, Ukrainisch

Tschechisch, Polnisch, Slowakisch, Obersorbisch

Bulgarisch, Mazedonisch, Kroatisch (jekavisch), Serbisch (ekavisch), Slowenisch

II. „Master i Margarita“ in 11 slawischen Sprachen (wie oben außer Obersorbisch)

Linguistische Randbedingungen:

Wortdefinition: basierend auf orthographischen Kriterien

Automatische Bestimmung der Textlänge

Teplé jarné slniečko práve zapadalo, keď sa pri Patriarchových rybníkoch zjavili dvaja občania. Prvý **z** nich – štyridsiatnik **v** sivom letnom obleku – bol nízky, zavalitý brunet **s** plešinkou, stľapčený klobúčik držal **v** ruke **a** hladko vyholenú tvár mu zdobili ozrutné okuliare **s** čiernym kosteným rámom. Druhý – plecnatý, ryšavkastý **a** strapatý mládenec **s** pepitovou čiapkou, zacapenou von **z** čela – mal na sebe károvanú rozhalenku, dokrkvané biele nohavice **a** čierne mokašiny.

Jede Kette von Segmenten zwischen zwei Leerzeichen = 1 Token

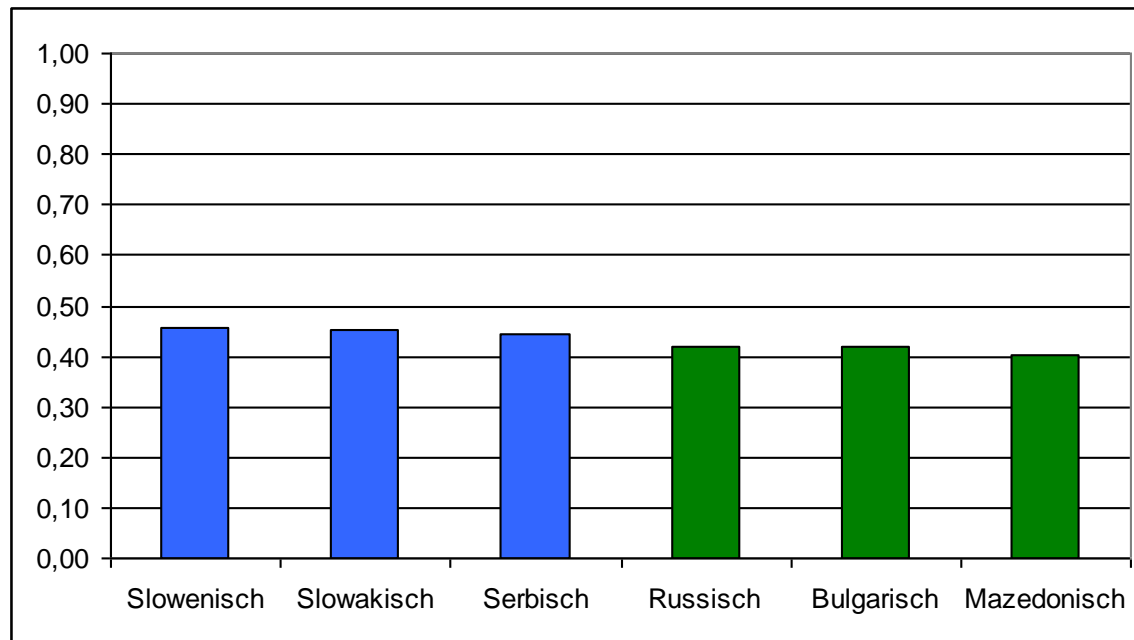
Keine Berücksichtigung der Frequenz:
Wortformen-Types (fett markiert)
= 62 Types

N	Word	Freq.	%	Lemmas
1	A	3	4,41	
2	S	3	4,41	
3	V	2	2,94	
4	Z	2	2,94	
5	BIELE	1	1,47	
6	BOL	1	1,47	
7	BRUNET	1	1,47	
8	ČELA	1	1,47	
9	ČIAPKOU	1	1,47	
10	ČIERNE	1	1,47	
11	ČIERNYM	1	1,47	
12	DOKRKVANÉ	1	1,47	
13	DRUHÝ	1	1,47	
14	DRŽAL	1	1,47	
15	DVAJA	1	1,47	

Wortlänge in Morphemen

Krupa/Altmann (1965): Intervall $<0,1>$

Sprache	Types	Morpheme	WO/MO
Slowenisch	574	1261	0,46
Slowakisch	486	1072	0,45
Serbisch	537	1209	0,44
Russisch	480	1151	0,42
Bulgarisch	520	1246	0,42
Mazedonisch	517	1282	0,40



“KZS” : Wortlänge (Types) 1. Kapitel

Sprache	Types	Morpheme	Wortlänge	<i>s</i>
Slowenisch	574	1261	2,20	0,94
Slowakisch	486	1072	2,21	1,08
Serbisch	537	1209	2,25	0,91
Russisch	480	1151	2,40	1,08
Bulgarisch	520	1246	2,40	1,01
Mazedonisch	517	1282	2,48	1,05

“MiM” : Wortlänge (Types) 1. Kapitel

Sprache	Types	Morpheme	Wortlänge	<i>s</i>
Serbisch	444	997	2,25	0,97
Slowenisch	436	991	2,27	1,11
Slowakisch	406	940	2,32	1,09
Bulgarisch	446	1117	2,50	1,11
Russisch	402	1031	2,56	1,33
Mazedonisch	449	1175	2,62	1,17

Test auf signifikante Unterschiede

Keine Normalverteilung der Wortlängenhäufigkeiten

Kruskal-Wallis-Test (nichtparametrisch)

Wortdefinition: unterschiedliche Ansätze

Kriterien:

- orthographisch
- phonetisch
- phonologisch
- morphologisch
- ...

(A) und (S) aus Sicht der theoretischen Linguistik

Strömungstheorie, Drift:

- Abbau synthetischer Konstruktionen und Ausbau analytischer Konstruktionen
(Sapir 1921, Vennemann 1984)

Kritik:

- gilt (wenn überhaupt) nur für indoeuropäische Sprachen
- für das Deutsche kein empirisches Nachweis eine „unidirektionalen“
Entwicklung (Roelcke 2002)
- Sprachen bewegen sich mehr oder weniger um einen bestimmten Grad an (A)
und (S), wie z.B. in slawischen Sprachen

Warum überschreitet Grammatik ständig die Grenzen zwischen Morphologie und Syntax?

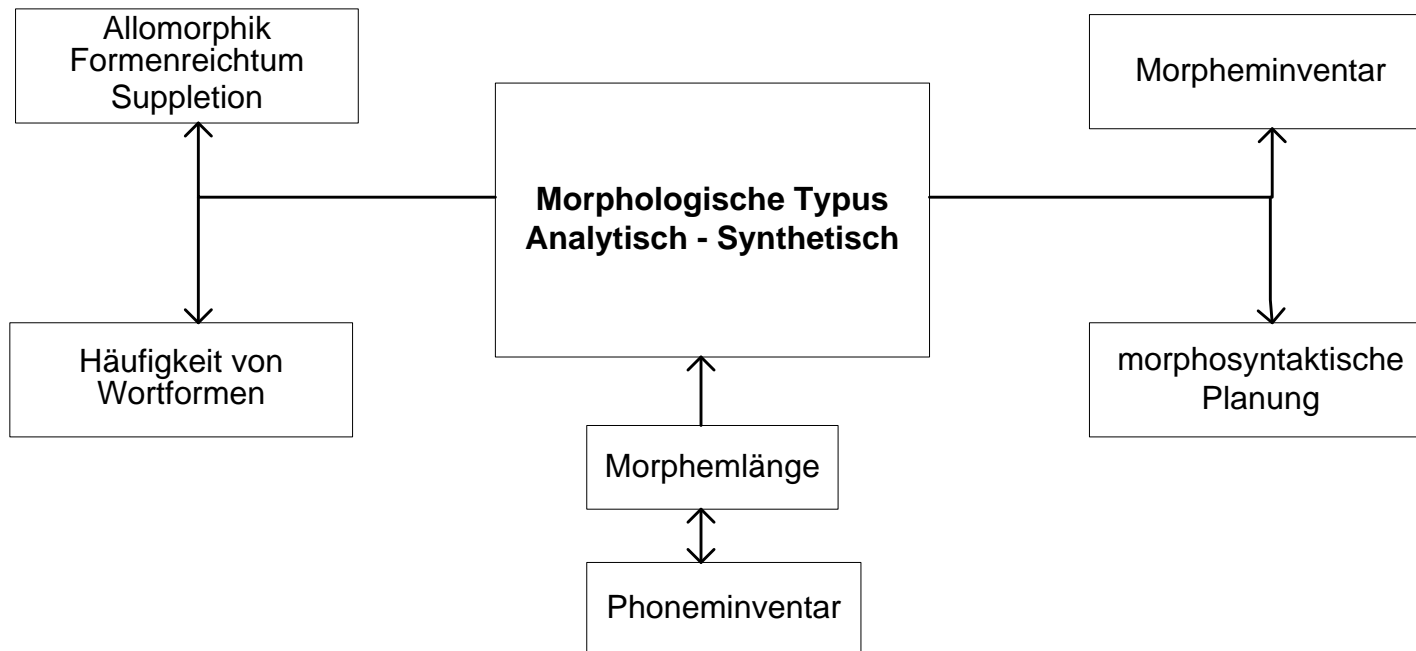
- Annahme von „Serialisierungsgesetzen“: Restriktionen in der Wortreihenfolge (Bybee 1985) und der Wortlänge (Leiss 1995)
- soziolinguistische Aspekte (Hinrichs 2000): (A) und (S) in Balkansprachen als Ausdruck von extremen Bedingungen in der mündlichen Sprachverarbeitung
- sprachökonomisch motivierte Anpassungs-Phänomene an verlangsamte Dekodierungs-Prozesse
- kommunikativer Zwang zur Vereinfachung des Sprachsystems und ein effektiveres Sprechverstehen
- analytische Struktur als ein entscheidender Kommunikationsvorteil beim gemischtsprachlichen (De)-Kodieren
- Sprachen als “selbstregulierende dynamische Systeme“, deren ständiger 'Antrieb' auf eine inhärente Antinomie zurückzuführen ist (Geisler 1982: 21)
- Analytisch/Synthetische Verfahren in einem „internen Konkurrenzkampf“
- alternative Kodierungsverfahren auf synchroner Ebene vs. diachrone Entwicklung

Languages can be regarded as homeostatic (or "self-regulating") systems, the state of a language at any one time being ‚regulated‘ by two opposing principles: The first of these (which is sometimes referred as the principle of least effort) is the tendency to maximize the efficiency of the system (...): its effect is to bring the syntagmatic length of words and utterances closer to a theoretical ideal. The other principle is ‚the desire to be understood‘: this inhibits the shortening effect of "least effort" by introducing redundancy at various levels“. (Lyons 1968: 90)

Sprachen als “selbstregulierende dynamische Systeme“, deren ständiger 'Antrieb' auf eine inhärente Antinomie zurückzuführen ist (Geisler 1982: 21)

- Analytisch/Synthetische Verfahren in einem „internen Konkurrenzkampf“
 - beobachtbar auf synchroner Ebene (alternative Kodierungsverfahren)
 - beobachtbar auf diachroner Ebene (keine Aussage über die Richtung der Entwicklung)

Frage von Zusammenhängen und Wechselbeziehungen



Ronneberger-Sibold, Elke (1980): Sprachverwendung, Sprachsystem. Ökonomie und Wandel. Tübingen: Niemeyer (Linguistische Arbeiten, 87).

weitere Faktoren:

- Silbenstruktur
- Textlänge (insbesondere bei Parallel-Texten)
- Grad an Polysemie/Synonymie
- morphosyntaktische Organisation (Wortstellung, Satzstruktur, Satzlänge)