

**Emmerich Kelih
(Graz)**



Аналитизм/Синтетизм в славянских языках:

– Параллельный корпус как основа анализа

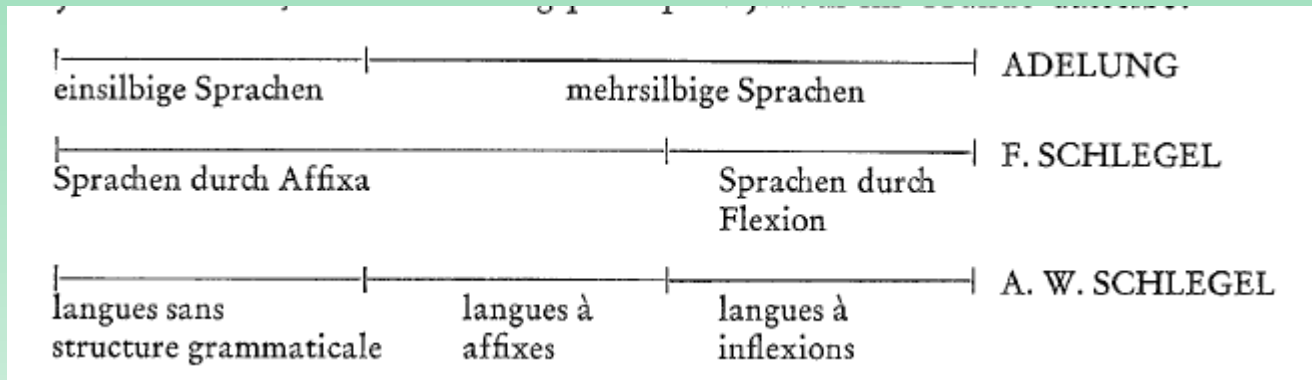
- Institut für Slawistik, Universität Graz
- <http://www-gewi.uni-graz.at/quanta/> [Graz-project on Quantitative Text-Analysis]
- <http://www.uni-graz.at/emmerich.kelih/> [emmerich.kelih@uni-graz.at]

Цель данного исследования

Морфологическая классификация славянских языков

1. Аналитический (A) и синтетический (S) типы: лингвистическое значение термина
2. „state of the art“ в славистике
3. Параллельный корпус как надежный источник исследования A/S?
4. Как можно эмпирически определить степень A/S?
5. эмпирические результаты длины слов в 12 славянских языков
6. Итоги и перспективы

Немного истории: Морфологическая классификация



Дальнейшее развитие классификации:

- изолирующие, агглютинативные, инкорпорирующие, флективные (фузионные) ЯЗЫКИ
- морфологически „оформленный“ vs. „неоформленный“
- аналитический vs. синтетический типы

„Jene äussere Sprachform, und mithin auch die innere, ist eine **analytische**, das heisst der **Gedanke wird in seine Bestandteile** zerlegt und in diesem **zerlegten Zustande zum Ausdrucke gebracht.**“

„Der Analyse entspricht als (**synthetisches**) **Ergebniss** ein **organisch gegliederter Körper**, das heisst ein Satz oder ein Satzwort, worinnen das Ganze und die Theile zu einander in Wechselwirkung stehen.“

„So ergeben sich **zwei einander nothwendig ergänzende grammatische Systeme**: das eine nenne ich das **analytische**, weil in ihm die **Spracherscheinungen durch Zerlegung** erklärt werden, das andere nenne ich das **synthetische**, weil es lehrt die **grammatischen Mittel** zum Aufbaue der Rede zu verwerthen.“

(von Gabelentz 1901: 81-85)

Современное понимание

„In der heutigen Typologie befaßt man sich kaum mehr mit der klassischen, aus der Frühzeit der Romantik stammenden Fragestellung einer Einteilung in „geformte“ und „formlose“, oder in „isolierende/agglutinierende/flektierende/polysynthetische“ Sprachen.“

„Nur gelegentlich wird die alte, auf August Wilhelm Schlegel zurückgehende Diskussion über den „analytischen“ Sprachbau aufgegriffen und mit neuen Einsichten weitergeführt.“

Bossong, G. (2001)

Мнение славистов ...

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

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.

дезидераты по Hinrichs-у (2000)

1. уточнение термина аналитический/синтетический
2. различные типы аналитизма
3. выяснение теоретического статуса классификации A/S
4. когнитивная интерпретация A/S
5. изменение статуса A/S во времени (диахрония)
6. эмпирический анализ A/S
7. сравнение степени A/S в различных языках

Уточнение термина A/S

Аналитический тип языка:

выражение грамматических значений отдельными служебными словами;
перераспределение грамматической информации с морфологического на синтаксический уровень

... или:

выражение грамматической информации как правило самостоятельными словоформами
(Hinrichs (2004a: 21)

... или:

появление многокомпонентной конструкции вместо «простой» флективной конструкции ...
(Reiter 2002: 87)

Лингвистические следствия:

1. дефиниция слова
2. вопрос о пограничных сигналах
3. когнитивное значение слова
4. A/S в сравнительных исследованиях

- из-за отсутствия универсального определения слова и словоформы нужно полностью отказаться от классификации на аналитические и синтетические языки (Schwegler 1990: 46)
- абсолютная значимость границ слова является теоретическим конструктом языковой типологии (Haarman 2004: 73)
- „Much that has been written about the word is decidedly eurocentric“.
(Dixon/Aikhenvald 2002: 2)

Потенциальные признаки A/S

- (1) распад/исчезновение падежной системы (включая изменения падежной системы, повышение синкретизма и.т.д.)
- (2) исчезновение согласования
- (3) десемантизация предлогов
- (4) усиление предложной системы
- (5) выражение сравнительных форм аналитическим/синтетическим способом
- (6) устройство временной системы: образование формы будущего времени ...
- (7) различные явления в словообразовании

(Hinrichs 2000a: 92)

... ..

Примеры из различных славянских языков:

выражение сравнительных форм в словенском языке:

I. Синтетический тип:

положительная - сравнительная – превосходная степень

lep – lepši – **najlepši**

težak – težji - **najtežji**

bogat – bogatejši – **najbogatejši**

II. аналитический тип

(у многосложных словах, с окончанием на --en, -av, -ast, -at и слова обозначающие
цвет

zaželen – bolj zaželen – najbolj zaželen

duševen – bolj duševen – najbolj duševen

muhast – bolj muhast – najbolj muhast

Образование формы будущего времени в сербском

1. аналитический тип

краткая форма от хтети + инфинитив

ћу ћемо

ћеш ћете + инфинитив радити, питати

ће ће

2. инфинитив – ти + краткая форма от хтети

видећемо, питаћу, радићете

[но: доћи ћу]

3. аналитический тип: краткая форма хтети + да-конструкция + настоящее время

то ћемо да видимо

и он ће да доће

студенти неће да се врате до јесени ...

→ аналитический + синтетический типы
образования

→ различные типы аналитизма

Русский: анти-аналитический язык?

- высокая частота предложных конструкций

операция по пересадке сердца

план по выпуску изделий широкого употребления vs. план выпуска

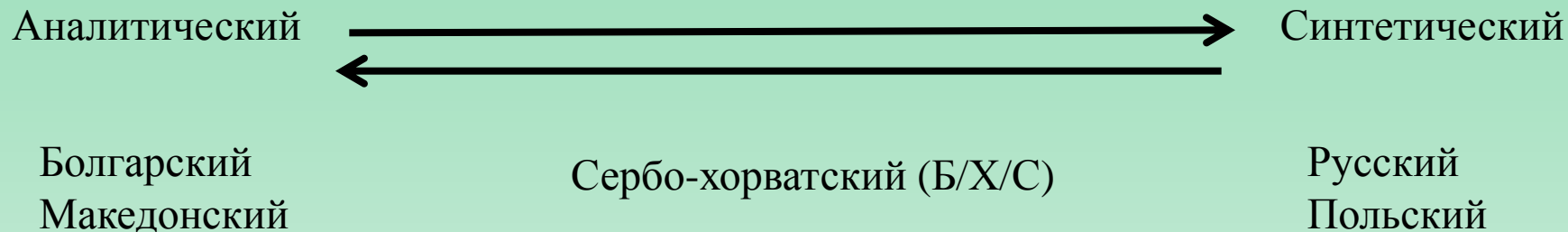
инженер по акустике vs. *akustični инженер

учебник по математике vs. учебник математики

„... eine schon fast parasitäre Polysemierung dieser Universalpräposition ... bei der der potentielle Zuwachs an Analytismus einen gering zu veranschlagenden Nebeneffekt darstellt“ (Weiss 2004: 269)

- высокая частота аббревиатур
- тенденция к образованию композитов

Где находятся на А/С славянские языки ?



В славистике считают, что болгарский и македонский языки обладают высокой степени аналитичности, а русский и польский считаются очень синтетическим. Это общее место в славистике разделяемое многими учеными.

Между этими полюсами располагается сербохорватский язык как типологически смешанный язык. Он синтетически по морфологии имени, а аналитически по морфологии глагола. (Hinrichs 2000: 91)

Открытый вопрос

Какие данные, свойства и параметры лежат в основе такой «классификации»?

Выводы:

- Аналитический/синтетический тип как градуальный и относительный концепт
- A/S подлежит квантификации и квантитативному анализу
- используемый материал должен обеспечить сравнимость результатов.

→ Эмпирический материал: Корпус параллельных текстов
«Как закалялась сталь (KZS) в 12 славянских языках

→ Длина слова в параллельных текстах = индикатор A/S (Greenberg 1960)

длина слова (в фонемах, графемах, слогах)

число морфем на слово = индикатор A/S Greenberg (1960: 187)

Характеристики KZS

- Текст: “Как закалялась сталь” (KZS)

- роман написаны в 30-их гг. XX в.

- автор: Н.А. Островский

- 10 глав в 12 языках:

белорусский, русский, украинский

чешский, польский, словацкий, верхнелужицкий

болгарский, македонский, хорватский (ijekavica), сербский (ekavica), словенский

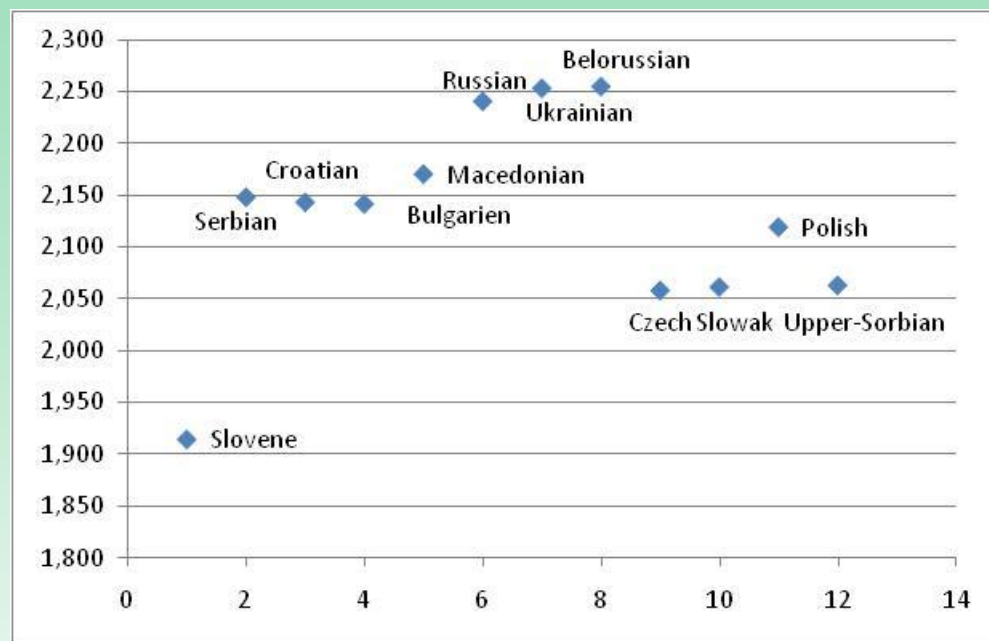
→ „Input“: графическое слово

→ длина слова (WL): измеряется числом слогов
длина словоупотреблений (Tokens)
= синтагматический уровень

→ измерение осуществляется автоматический

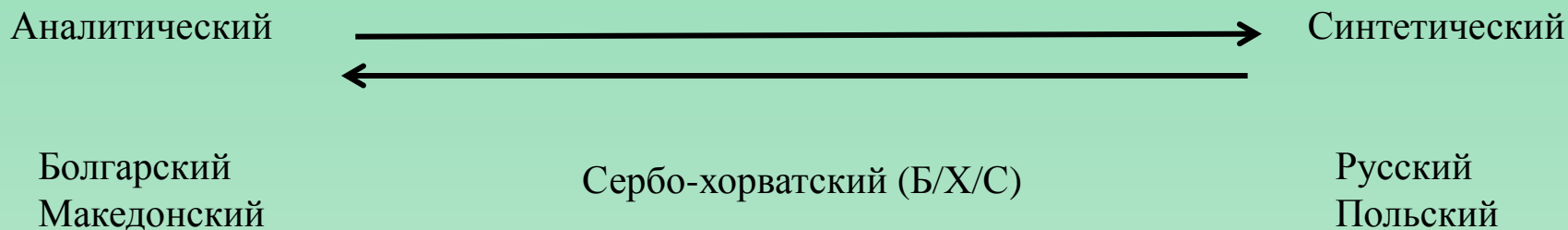
“Как закалялась сталь” : длина слов (словоупотреблений)

Nr.	language	Tokens	WL/Syl.
1	Slovene	62655	1,914
2	Serbian	56230	2,148
3	Croatian	56424	2,143
4	Bulgarian	57174	2,142
5	Macedonian	58837	2,171
6	Russian	49675	2,241
7	Ukrainian	49612	2,254
8	Belorussian	50010	2,255
9	Czech	52180	2,058
10	Slovak	52099	2,062
11	Polish	52737	2,119
12	Sorbian	58484	2,063



→ Чем длиннее «слово», тем более язык синтетический!

→ Получается типологический ряд славянских языков!



Индикатор A/S в славянских языках:

аналитический синтетический

Slo.	CZ.	SK	Sorb.	Pl.	Bulg.	Cro.	Serb.	Mac.	Rus.	Ukr.	Belor.
1,91	2,06	2,06	2,06	2,12	2,14	2,14	2,15	2,17	2,24	2,25	2,26

на синтагматическом уровне

Предварительные результаты:

- Македонский/Болгарский вряд ли могут быть названы «высоко» аналитическими языками
- Словенский, чешский, словацкий, верхнелужицкий являются аналитическими языками
- Синтетические: восточнославянские языки
- Польский в анализируемом материале тяготеет к «синтетическим» языкам

Итоги

1. параллельные тексты как адекватная основа анализа A/S
2. длина «слова» может быть использована как индикатор A/S
3. различия в длине «слов» фактически минимальные, но зато лингвистически объяснимые

Перспективы:

1. анализ дополнительных текстов
2. разработка альтернативных индикаторов (лексический уровень: закон Ципфа, «*hax legomena*»)

Литература (выбор)

Bossong, Georg (2001): Die Anfänge typologischen Denkens im europäischen Rationalismus. In: Haspelmath, Martin; König, Ekkehard Oesterreicher Wulf; Raible, Wolfgang (Hg.): Language Typology and Language Universals/ Sprachtypologie und sprachliche Universalien. Berlin u.a.: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 20,1), S. 249–264.

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

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.

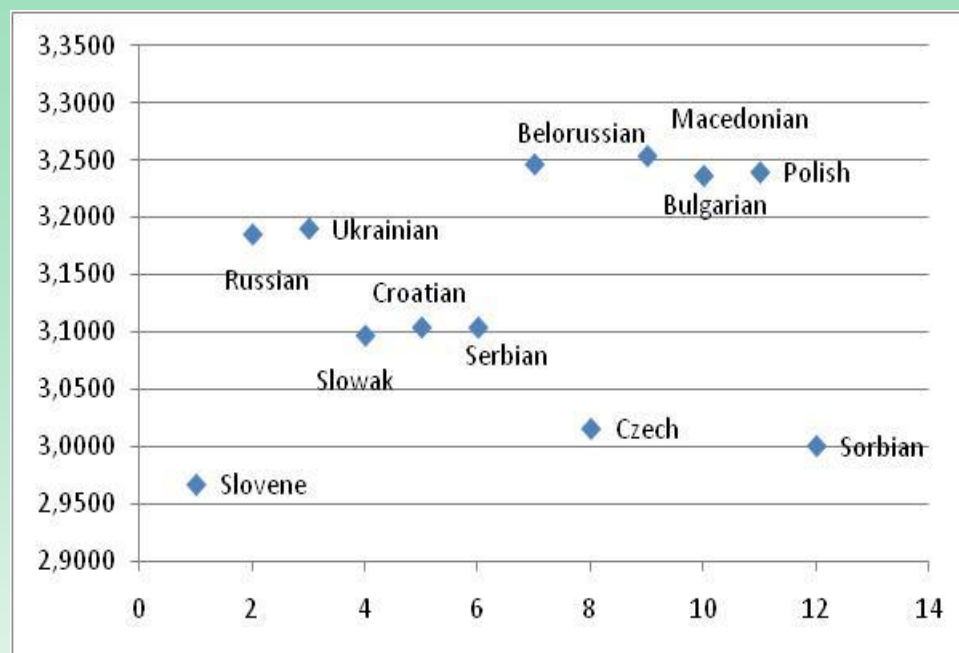
Hinrichs, Uwe; Büttner, Uwe (Hg.): Die Südosteuropa-Wissenschaften im neuen Jahrhundert: Akten der Tagung vom 16. - 19.10.1999 an der Universität Leipzig. Wiesbaden: Harrassowitz, S. 107–128.

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.

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

“Как закалялась сталь” : длинна типов (словоформ)

Nr.	language	Types	Wol/Syl
1	Slovene	13946	2,9667
2	Serbian	13642	3,1037
3	Croatian	13737	3,1038
4	Bulgarian	12308	3,2358
5	Macedonian	11465	3,2531
6	Russian	15053	3,1849
7	Ukrainian	14645	3,1899
8	Belorussian	14858	3,2459
9	Czech	14136	3,0155
10	Slovak	14027	3,0967
11	Polish	14978	3,2390
12	Sorbian	14574	3,0006

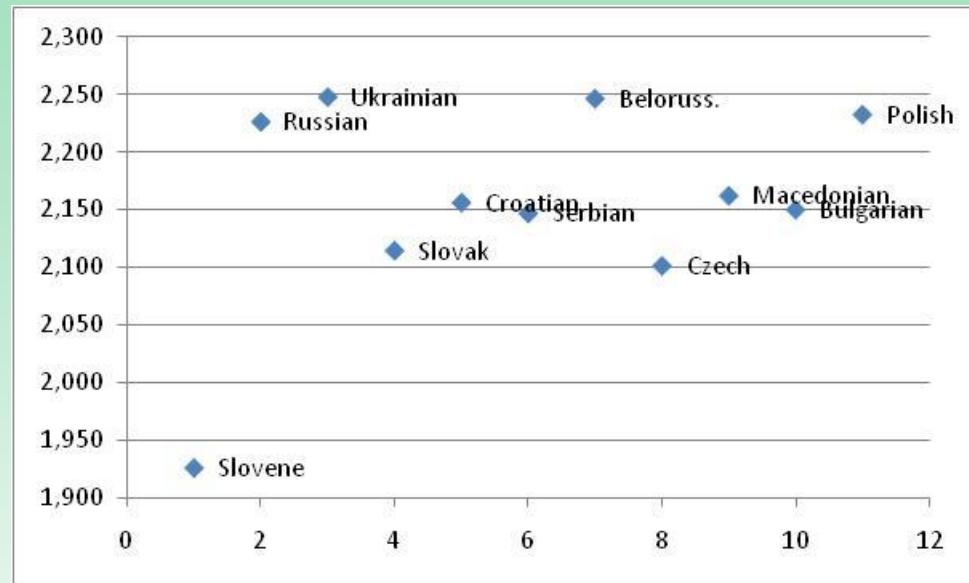


→ получается иной типологический ряд языков

→ самое замечательное: македонский + болгарский являются синтетическими языками

анализ дополнительного материала: KZS + Мастер и Маргарита = один корпус
WOL-Tokens

Slovene	194728	1,926
Russian	163426	2,227
Ukrainian	162345	2,248
Slovak	163789	2,115
Croatian	179964	2,156
Serbian	183648	2,147
Beloruss.	158815	2,247
Czech	160812	2,102
Macedonian	190852	2,162
Bulgarian	181554	2,150
Polish	172161	2,233



аналитический

синтетический



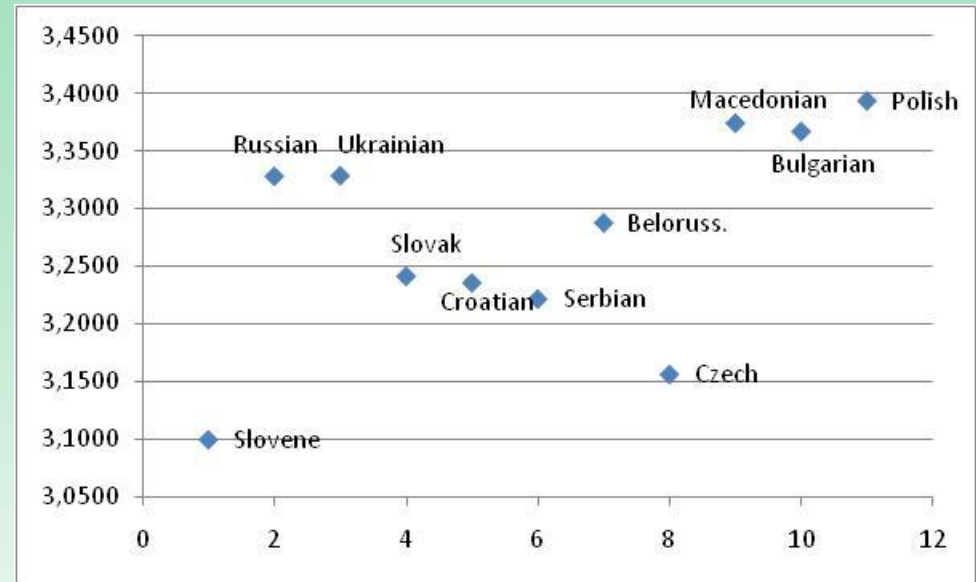
Slo.	Cz.	Sk.	Serb.	Bulg.	Cro.	Mac.	Rus.	Pol.	Belor.	Ukr.
1,93	2,10	2,11	2,15	2,15	2,16	2,16	2,23	2,23	2,25	2,25

Slo.	CZ.	SK	Sorb.	Pl.	Bulg.	Cro.	Serb.	Mac.	Rus.	Ukr.	Belor.
1,91	2,06	2,06	2,06	2,12	2,14	2,14	2,15	2,17	2,24	2,25	2,26

(KZS)

анализ дополнительного материала: KZS + Мастер и Маргарита = один корпус
WOL-Types

Slovene	28419	3,0999
Russian	32874	3,3284
Ukrainian	34256	3,3289
Slovak	30885	3,2417
Croatian	29859	3,2359
Serbian	29851	3,2219
Beloruss.	30819	3,2880
Czech	31714	3,1565
Macedonian	24013	3,3745
Bulgarian	25761	3,3672
Polish	32711	3,3938



аналитический

синтетический

—————→

Slo.	Cz.	Serb.	Cro.	Sk.	Belor.	Rus.	Ukr.	Bulg.	Mac.	Pol.
3,10	3,16	3,22	3,24	3,24	3,29	3,33	3,33	3,37	3,37	3,39

Slo.	Sorb.	Cz.	Sk.	Serb.	Cro.	Russ.	Ukr.	Bulg.	Pol.	Belor.	Mac.
2,97	3,00	3,02	3,10	3,10	3,10	3,18	3,19	3,24	3,24	3,25	3,25

KZS

Длинна слов = число слогов/число словоформ

Длинна слов = число слогов / число Token (словоформ)

→ длинна текста играет важную роль

Тем длиннее (параллельный) текст, тем выше степень аналитизма.

Лингвистические проблемы использования параллельных текстов:

«Упрощение» (на синтаксическом уровне: на пример длина предложения уместается в переводимых текстах, разные лексические упрощения ...)

«Тяготение переводчика к экспликации»: более подробное изложение смысла/содержание переводчиком

отсутствие перевода некоторых частей текста (обнаруживается в время alignment)

Орфографическая словоформа/словоупотребление

Автоматический анализ Type/Token

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, stĺpčený 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 mokasíny.

Цепочка знаков между пробелами = 1

Token

→ 68 Tokens (словоупотреблений)

→ 62 Types (словоформ)

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	

Резюме (предварительные результаты)

Индикатор A/S в славянских языках:

аналитический На синтагматическом уровне синтетический



Slo.	CZ.	SK	Sorb.	Pl.	Bulg.	Cro.	Serb.	Mac.	Rus.	Ukr.	Belor.
1,91	2,06	2,06	2,06	2,12	2,14	2,14	2,15	2,17	2,24	2,25	2,26

На парадигматическом уровне



Slo.	Sorb.	Cz.	Sk.	Serb.	Cro.	Russ.	Ukr.	Bulg.	Pol.	Belor.	Mac.
2,97	3,00	3,02	3,10	3,10	3,10	3,18	3,19	3,24	3,24	3,25	3,25

analytic

synthetic

Syllable structure

Word Length

Morpheminventar

Flektional productivity

Degree of polysemy

text length (number
of words)

Degree of synonymy

Restrictive word
order

Zipf's a

Amount of hapax
legomena

Parts of speech
(absolute numbers)

h-point