

Корпусная типология

Самира Ферхеес

Типология 2021 – Семинар

Типология по корпусам

- ▶ Что делают типологи с корпусами?

Типология по корпусам

- ▶ Что делают типологи с корпусами?
- ▶ Смотрят частотности (как и все)
- ▶ Смотрят, как выражается X

Типология по корпусам

- ▶ Что делают типологи с корпусами?
- ▶ Смотрят частотности (как и все)
- ▶ Смотрят, как выражается X
- ▶ Wälchli (2007): intensional и extensional подходы к типологии
- ▶ Intension: семантическое определение → способы кодирования
- ▶ Extension: ситуация в контексте → способы кодирования

Типология по корпусам

Для сравнительных целей удобнее всего параллельные корпуса.



Multi-verb constructions

- (1) Ač-i-ne Man pat-ăm-a
 child-POSS3-DAT/ACC I.GEN to-POSS1SG-DAT
 il-se kil-ěr.
 take-CONV come-IMP2PL
 ‘bring him unto me’

чувашский (Wälchli 2007)

Multi-verb constructions

Table 1. Cross-linguistic diversity in multi-verb constructions.

		BRING	
		Verb solitarizing	Multi-verb constructions
RUN	Solit.	Dinka, Navajo, Russian	Ainu, Ewe, Khasi
	MVC	English, Guaraní, Maltese	Choctaw, Chuvash, Khoekhoe

(Wälchli 2007)

Multi-verb constructions

Wälchli (2007): BRING и RUN в Евангелии от Марка

BRING	RUN
1:32 <i>they brought unto him all that were diseased</i>	5:6 <i>he ran and worshipped him</i>
2:03 <i>bringing one sick of the palsy</i>	6:33 <i>and ran afoot thither out of all cities</i>
6:27 <i>and commanded his head to be brought</i>	6:55 <i>And ran through that whole region round about</i>
6:28 <i>And brought his head in a charger</i>	9:15 <i>and running to him saluted him</i>
7:32 <i>And they bring unto him one that was deaf</i>	10:17 <i>there came one running, and kneeled to him</i>
8:22 <i>and they bring a blind man unto him</i>	15:36 <i>And one ran and filled a sponge full of vinegar</i>
9:17 <i>I have brought unto thee my son</i>	
9:19 <i>bring him unto me</i>	
9:20 <i>And they brought him unto him</i>	
10:13 <i>And they brought young children to him</i>	
11:02 <i>and bring him</i>	
11:07 <i>And they brought the colt to Jesus,</i>	
12:15 <i>bring me a penny</i>	
12:16 <i>And they brought it</i>	
15:01 <i>and carried him away</i>	
15:16 <i>And the soldiers led him away into the hall</i>	
15:20 <i>and led him out to crucify him</i>	
15:22 <i>And they bring him unto the place Golgotha</i>	

Multi-verb constructions

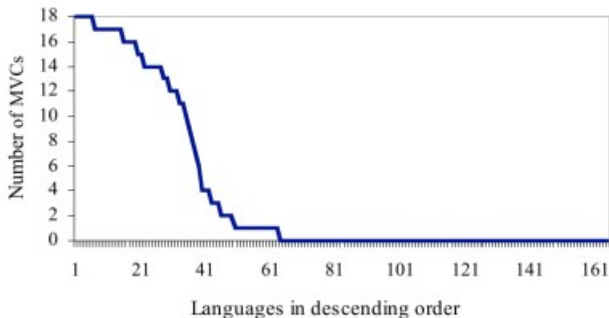
Использует ли язык MVC хоть один раз в каком-то домене

		BRING	
		Solit.	MVC
RUN	Solit.	65	12
	MVC	46	42

Multi-verb constructions

MVC в домене BRING образуют континуум

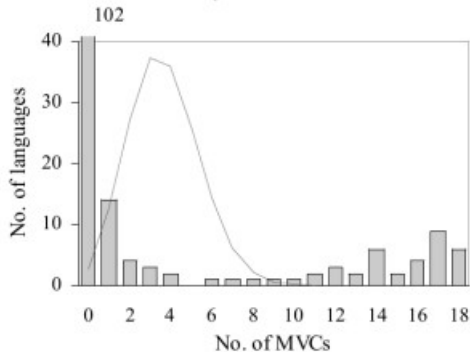
Figure 1. Number of MVC per language in the BRING domain (languages are ordered in descending order of the number of MVC).



Multi-verb constructions

Но дистрибуция не рандомная

Figure 2. Bipolar structure of the BRING domain (the line shows expected frequencies, the bars show the actual data).



Multi-verb constructions

Какая может быть причина вариативности внутри
языка и между языками?

Multi-verb constructions

Table 5. Number of languages according to MVCs in the two domains BRING and RUN (deviation from statistical expectation in brackets).

		BRING		
		0-1 (Low)	2-8 (Intermed.)	9-18 (High)
RUN	0 (Low)	70 [+15.9]	3 [-3.0]	4 [-12.8]
	1-3 (Intermed.)	36 [-2.0]	6 [+1.7]	12 [+0.2]
	4-6 (High)	10 [-13.9]	4 [+1.3]	20 [+12.6]

Multi-verb constructions

- ▶ В случае RUN больше промежуточных языков (33% напротив 8%)
- ▶ Ареальность: MVCs in BRING cluster strongly at various places in the Old World: West Africa (including Haitian Creole and Sranan), South-East, East, and South Asia, and Eastern New Guinea.
- ▶ Промежуточные языки встречаются на границе между кластерами языков, которые ближе к полюсам (но к сожалению Вэлхли не показывает где что)

Плюсы параллельных текстов

- ▶ Сравнимый материал для разных языков
- ▶ Дает представление о частотности (хотя бы в одном конкретном контексте), напр. *exceed*-компаративы в древнегреческом, латыни, и тибетском ([Wälchli 2007](#))

Минусы параллельных текстов

- ▶ Практические проблемы как отсутствие аннотации
- ▶ Отсутствие нужного контекста в доступном материале
- ▶ Если изучать не только европейские языки, то есть только Библия
- ▶ И ее нет для многих изолятов и вымирающих языков, и преобладают языки с письменностью

Грамматика притворства

Майсак (2021):

- ▶ Как выражаются значения типа [X ведет себя как будто Y] в нахско-дагестанских (и соседних тюркских) языках?
- ▶ В типологии обсуждается в двух контекстах: complement clauses и сравнительные конструкции

Грамматика притворства

- ▶ Евангелие от Луки
24:28 *Исус сделал вид, что Он собирается идти дальше.*
- ▶ В 16 разных языках (включая 2 бесписьменных)

Грамматика притворства

- (2) Жи-в колъи=ло седи
сам-М[ABS] немного=ADD вперед
в-улин-ну в-укку-дя=**ГЪОДДУ**
М-уходить-INF М-должен-FUT=QUOT
ххвел игъи-дду ГЪО-Ш-ди.
притворство[ABS] делать-PRF ЭТОТ-OBL.М-ERG
'(что) Он должен идти дальше вперед, мол,
Он притворился'
андийский (**Майсак 2021: 231**)

Грамматика притворства

- (3) Жи-в дагъа-в=ги це<в>ехун ине
сам-м[ABS] еще-м=ADD вперед<м> идти.INF
кк-ол-е-б ххвел
случаться-PRS-PTCP-N притворство[ABS]
гъабу-на гъе-с.
делать-AOR ЭТОТ-ERG.М
'(что) Он должен идти дальше вперед, Он
притворился'
аварский (Майсак 2021: 230)

Грамматика притворства

- (4) abdulla xoroša=ɬo w-ux-e
abdulla sheperd=FUNC M-stop-NAV
'Абдулла притворяется чабаном.'
андийский: зиловский
- (5) mak'i k'wahaɬal=talu /=ɬun w-ix-ata
child sick=QUOT =FUNC M-stop-PROG.CVB
ida
cop
'Ребенок притворяется больным.'
ботлихский

(Мои полевые данные)

Переводы Библии

- ▶ hagiolect effects
- ▶ Использование определенного регистра, калькирование из более престижного языка, лексический конкорданс
- ▶ Разные стратегии перевода → разные эффекты (De Vries 2007)

Переводы Библии

- ▶ hagiolect effects
- ▶ Использование определенного регистра, калькирование из более престижного языка, лексический конкорданс
- ▶ Разные стратегии перевода → разные эффекты (De Vries 2007)
- ▶ Не очень понятно, какой агиолект мы можем ожидать в переводе Библии от носителя бесписьменного дагестанского языка

Альтернативы?

Альтернативы?

Universal Dependencies

Current UD Languages

Information about language families (and genera for families with multiple branches) is mostly taken from [WALS Online](#) (IE = Indo-European).

1	Abaza	1	<1K	🗨️	Northwest Caucasian
2	Afrikaans	1	48K	🗨️🗨️	IE, Germanic
3	Akkadian	2	25K	🗨️🗨️	Afro-Asiatic, Semitic
4	Akunse	1	<1K	🗨️🗨️	Tupian, Tupari
5	Albanian	1	<1K	🗨️	IE, Albanian
6	Amharic	1	10K	🗨️🗨️🗨️	Afro-Asiatic, Semitic
7	Ancient Greek	2	416K	🗨️🗨️	IE, Greek
8	Apurina	1	<1K	🗨️🗨️	Arawakan
9	Arabic	3	1,042K	🗨️🗨️	Afro-Asiatic, Semitic
10	Armenian	1	52K	🗨️🗨️🗨️	IE, Armenian

Корпуса синтаксических структур в разных языках (100+); разные жанры текстов + единая схема аннотации синтаксических связей

Альтернативы?

Multi-Cast

Multi-CAST

Multilingual Corpus of Annotated Spoken Texts

[Annotations](#)[The corpora](#) +[Research](#)[Contribute](#)[People](#)[More](#) +

Multi-CAST, the *Multilingual Corpus of Annotated Spoken Texts*, is a collection of annotated texts from a typologically diverse section of languages.

Полевые записи на разных языках + единая схема аннотации для кодирования актантов и индексации референтов; языков пока мало и только монологи

Корпусная типология

- ▶ Что делают типологи с корпусами?
- ▶ Смотрят частотности (как и все)
- ▶ Смотрят, как выражается X
- ▶ ! Сравнивают морфологическую сложность языков и проверяют старые гипотезы (напр. универсалии) (Levshina 2021: 9–11)

Abbreviations

1	first person 6
2	second person 6
3	third person 6
ABS	absolutive 19, 20
ACC	accusative 6
ADD	additive 19, 20
AOR	aorist 20
CONV	converb 6
COP	copula 21
CVB	converb 21
DAT	dative 6
ERG	ergative 19, 20
FUNC	functionive 21
FUT	future 19
GEN	genitive 6
HAB	habitual 21
IMP	imperative 6
INF	infinitive 19, 20
M	masculine 19–21
N	neuter 20
OBL	oblique 19
PL	plural 6
POSS	possessive 6
PRF	perfect 19
PROG	progressive 21
PRS	present 20
PTCP	participle 20
QUOT	quotative 19, 21
SG	singular 6

Литература I



De Vries, Lourens. 2007. Some remarks on the use of Bible translations as parallel texts in linguistic research. *STUF. Language Typology and Universals* 60(2). 148–157.



Levshina, Natalia. 2021. Corpus-based typology: applications, challenges and some solutions. *Linguistic Typology* (aop). https://pure.mpg.de/rest/items/item_3289556_3/component/file_3310409/content.

Литература II



Wälchli, Bernhard. 2007. Advantages and disadvantages of using parallel texts in typological investigations. *STUF. Language typology and universals* 60(2). 118–134.



Майсак, Тимур А. 2021. Грамматика притворства: к внутригенетической микротипологии одной конструкции в нахско-дагестанских языках. In Тимур А. Майсак, Нина Р. Сумбатова & Яков Г. Тестелец (eds.), *Дурхъаси хазна. сборник статей к 60-летию р.о. муталова*, 220–257. Москва: Буки Веди.