Корпусная типология

Самира Ферхеес

Семинар по типологии 2021/2022

Слайды: github.com/sverhees/site
> teaching > typology 2021/2022

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 (грамматики)
- Extension: ситуация в контексте → способы кодирования (параллельные тексты)

Типология по корпусам

Для сравнительных целей удобнее всего параллельные корпуса.

Multi-verb constructions

- Wälchli (2007) предикаты из одного глагола напротив multi-verb constructions в параллельных текстах
- BRING и RUN
- Евангелие от Марка
- 165 языков, ареально разнообразная выборка

Multi-verb constructions

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

чувашский (Wälchli 2007)

Но также, например, *come running* на английском.

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

Multi-verb constructions

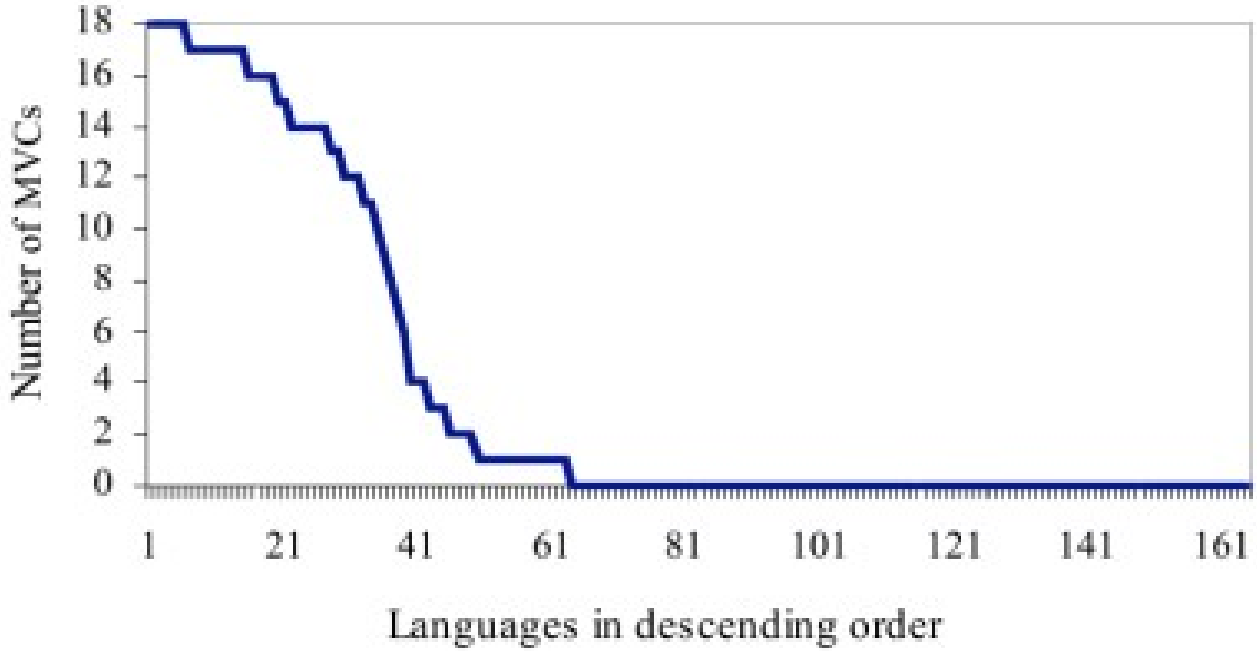
Использует ли язык MVC хоть один раз в каком-то домене

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

Multi-verb constructions

MVC в домене BRING образуют континуум

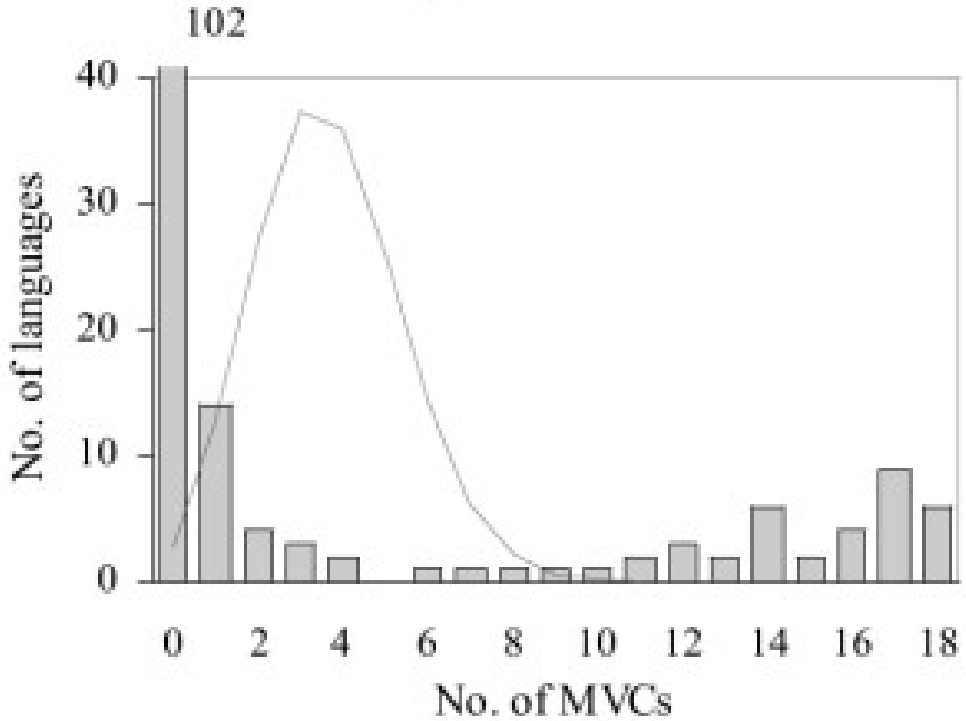
Figure 1. Number of MVC per language in the BRING domain (languages are ordered in descending order of the number of MVC).



Multi-verb constructions

Но диструбиция 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%)
- В целом BRING более показателен
- Напр. в плане ареальности: 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.
- Какие для этого могут быть причины?

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 У] в нахско-дагестанских (и соседних тюркских) языках?
- В типологии обсуждается в двух контекстах: complement clauses и сравнительные конструкции, ср. *X притворяется, что*, *X делает вид, что*, *X pretends that*, *X acts like*

Грамматика притворства

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

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 игьи-дду гьо-ш-ди.
притворство[ABS] делать-PRF этот-OBL.M-ERG
‘(что) Он должен идти дальше вперед, мол,
Он притворился’

андийский (Майсак 2021: 231)

Грамматика притворства

(3) Жи-в дагьа-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]
гьабу-на гье-с.
делать-AOR этот-ERG.М
‘(что) Он должен идти дальше вперед, Он притворился’

аварский (Майсак 2021: 230)

Грамматика притворства

(4) abdulla χoroša=ɬo w-uʁ-e
abdulla sheperd=FUNC M-stop-HAB
‘Абдулла притворяется чабаном.’
андийский: зиловский

(5) makʼi kʼʷaħal=talu /=ɬun w-iʁ-ata
child sick=QUOT =FUNC M-stop-PROG.CVB
ida
COP
‘Ребенок притворяется больным.’
ботлихский

(Мои полевые данные)

Грамматика притворства

- Пример грамматики притворства показывает, что бесписьменный язык андийский использует некоторый “литературный” оборот аварского
- (Скорее всего не увидев аварский перевод)
- Но при этом сохраняет часть “обычную” конструкцию, у которой нет эквивалента в аварском

Переводы Библии

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

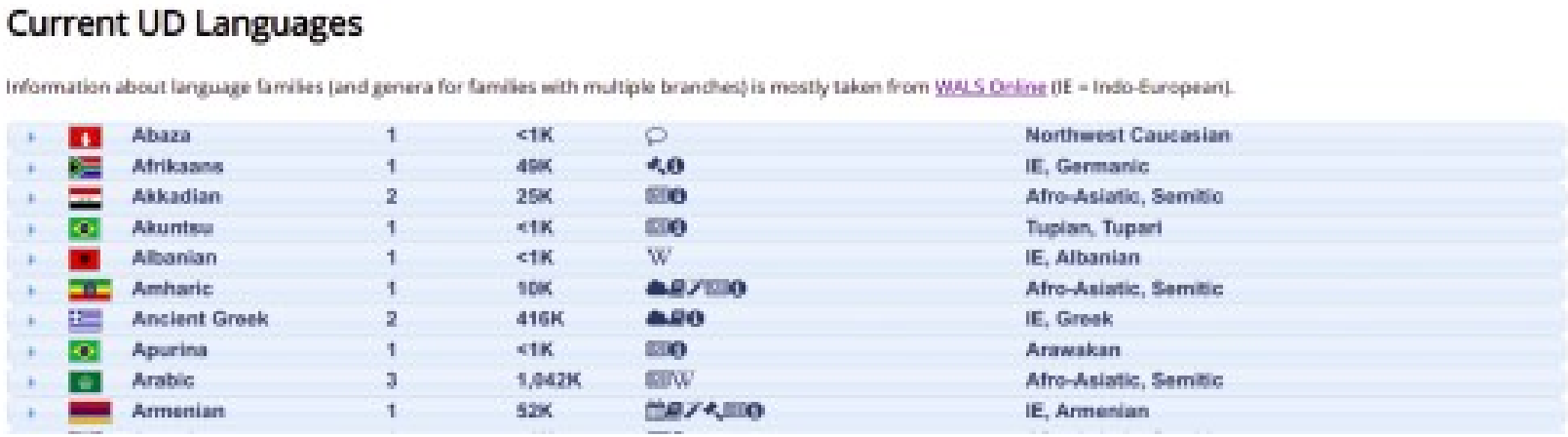
Переводы Библии

- hagiolect effects
- Использование определенного регистра, калькирование из более престижного языка, лексический конкорданс
- Разные стратегии перевода $\rightarrow$ 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).

Abaza	1	<1K		Northwest Caucasian
Afrikaans	1	49K		IE, Germanic
Akkadian	2	25K		Afro-Asiatic, Semitic
Akuntsu	1	<1K		Tupian, Tuparí
Albanian	1	<1K		IE, Albanian
Amharic	1	10K		Afro-Asiatic, Semitic
Ancient Greek	2	416K		IE, Greek
Apurina	1	<1K		Arawakan
Arabic	3	1,042K		Afro-Asiatic, Semitic
Armenian	1	52K		IE, Armenian

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

Альтернативы?

Multi-Cast

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

Альтернативы?

- Можно и создать параллельный корпус переводов релевантных контекстов?

Альтернативы?

- Можно и создать параллельный корпус переводов релевантных контекстов?
- Используется, например, в московской школе лексической типологии
- В работе Dahl (1985) использовалась база заполненных анект про вид и время

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)
- Корпусам необязательно быть параллельными, чтобы быть сравнимыми (см. исследование аналитичности разных языков в Levshina (2021))

Abbreviations

1	first person 7
2	second person 7
3	third person 7
ABS	absolutive 20, 21
ACC	accusative 7
ADD	additive 20, 21
AOR	aorist 21
CONV	converb 7
COP	copula 22
CVB	converb 22
DAT	dative 7
ERG	ergative 20, 21
FUNC	functive 22
FUT	future 20
GEN	genitive 7
HAB	habitual 22
IMP	imperative 7
INF	infinitive 20, 21
M	masculine 20–22
N	neuter 21
OBL	oblique 20
PL	plural 7
POSS	possessive 7
PRF	perfect 20
PROG	progressive 22
PRS	present 21
PTCP	participle 21
QUOT	quotative 20, 22
SG	singular 7

Литература I

- Dahl, Östen. 1985. *Tense and aspect systems*. New York: Basil Blackwell.
- 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.