

Клитики в параллельном корпусе «Слова о полку Игореве»

А.В.Циммерлинг (Гос ИРЯ им.
А.С.Пушкина/МПГУ/Ияз РАН

fagraey64@hotmail.com

<https://pushkin.academia.edu/AntonZimmerling>

Talk venue

- ✓ Чтения памяти А. А. Зализняка. 21 декабря 2018, пятница, Институт языкознания.
- ✓ Проблемная группа «Логический анализ языка».

Факты и мешки

- *...пустые они не стоят.*
- Многие считают, что лингвистика не обладает предсказательной силой. И не является <естественной> наукой...
- Но некоторые умеют предсказывать распределение языковых фактов, не зная всех деталей заранее.
- Одни лингвисты умеют *объяснять*, строя *модели*, другие умеют точно *описывать* наблюдаемую картину.
- Не все в языке доступно непосредственному наблюдению. Идеал лингвистического руссоизма – отринуть все модели и теоретические понятия и получить удовлетворительное описание сложных объектов (языковых систем) на основе созерцания – недостижим.
- Ср. не растущие в «лесу наблюдений» модельные понятия акцентной парадигмы, граммемы, морфологического падежа, клитики, согласования, подлежащего, семантической роли, сферы действия, дерева предложения, темы, ремы.

Модели Зализняка

- Научное величие А.А.Зализняка — в том, что его модели всегда опираются на точные описания предметной области.
- Его описания настолько точны, что возникает иллюзия, что модели за ними вообще нет. Это высшая похвала в лингвистической парадигме старого типа и недостаток в глазах части современных догматиков.
- Точность всегда обеспечивалась или полным перебором релевантных фактов или умением выдвинуть правдоподобную гипотезу на основе рассмотренной части фактов.

Иллюстрация: Кирик и Зализняк

- Объяснение дистрибуции конструкции им.п. + инфинитив в истории русского языка [Зализняк 2004: 157].
- А.А.Зализняк по ходу отметил, что после инфинитивов, зависящих от глаголов *хотѣти*, *велѣти*, писатель Кирик (н. XII в.) употреблял не им.п., а вин.п.: *велить блѹсти недѣлю и соубоутоу* (К. 73), *повелѣ дати опитемью* (И. 41).
- На первый взгляд, это частное наблюдение над конкретным текстом. Скорее всего, ААЗ не проверял большое количество древнерусских и среднерусских текстов, кроме берестяных и пергаменных грамот.
- Однако Кирик — важнейший, *модельный* источник по синтаксису конструкции им.п. + инфинитив в русском языке домонгольского периода, его употребление показательно потому, что он оставил длинный связный текст.

Кирик и Зализняк (2)

- ААЗ правильно установил, что один и тот же автор употреблял и им.п., и вин.п. при инфинитиве и правильно предположил, что они не были распределены хаотично.
- Это решение имплицитно соответствует тому, что называется гипотезой о дифференцированном маркировании объекта [Aissen 2003; Ронько 2017], правда ААЗ не указывает фактора, стоящего за распределением и опускает научнообразные комментарии и термины вроде DOM, DAM. Sapiienti sat.

ААЗ и статистика

- Запрета на употребление им.п. после каузативных и волитивных глаголов в древнерусском и среднерусском действительно нет.
- Однако какова статистика? (ее ААЗ, видимо, не собирал).
- Проверим по среднерусскому корпусу НКРЯ http://ruscorpora.ru/search-mid_rus.html (его в 2004 не было).

Тенденция, *однако*

- Подсчет по 110 ИГ в позиции после инфинитива в среднерусском корпусе НКРЯ

Всего им.п.	Без Caus и Vol	Всего вин.п.	Без Caus и Vol
1056	967	1516	531
100%	92, 75%	100%	35 %

Модели упорядочения клитик по ААЗ

- ААЗ предложил две основные модели упорядочения клитик (на базе древнерусского языка) – [Зализняк 1993; 2008].
- 1. Правила контактного расположения клитик, соответствующие *грамматике рангов*, т.е. табличному алгоритму, выстраивающему любую пару клитик (а, b) в единственном порядке. В стандартном случае, описанном ААЗ, имеется однозначное соответствие между множеством клитик {с} и клеткой/столбцом таблицы, который члены этого множества могут замещать.
- 2. Правила реализации клитик вне позиции кластеризации, т.н. правила барьера [Зализняк 1993: 287].

Закон Ваккернагеля по ААЗ

- ААЗ доказал, что в языке с законом Ваккернагеля (например, древнерусском) правила обеих типов (правила рангов и правила барьера) сочетаются друг с другом. Правила барьера – не отклонение от ЗВ, как думали некоторые недальновидные слависты, а его частный случай [Зализняк 2008: 220]. Распределение контекстов с барьером и без барьера регулируется правилами.
- Значение этого открытия выходит далеко за рамки описания древнерусского языка. Это — общий принцип функционирования систем порядка слов с цепочками клитик: клитики отличны от морфем и сохраняют синтаксическую свободу.

Проверка правил рангов в историческом корпусе

- Древнерусский язык: Правило рангов клитик-частиц, *6 клеток.
- 25 засвидетельствованных цепочек.
- Среднерусский язык: Правило рангов клитик-частиц, 5 клеток.
- 442 засвидетельствованных цепочки.

Др.рус. частицы-энклитики

ЖЕ	ЛИ	БО	ТИ ₁	БЫ	ДЪ(И)
4					
		5			
			1		
				5	
	1				
				3	
					1
				5	

Ср.рус. частицы-клитки

ЖЕ	ЛИ	БО	*ТИ ₁	БЫ	ДЕ(И)
31					
		67			
				79	
					178
	1				
				4	
				4	
					1
					25
	Чтения памяти Андрея Анатольевича Зализняка. Институт языкознания РАН, 21.12.2018			51	

Славянские языки, клитики и статистика

- Не все в массиве / корпусе древних памятников дает убедительную статистику и дает почву для статистического анализа вообще.
- В современных славянских языках есть ограничения на совместную реализацию местоименных невозвратных клитик дат.п. и вин.п., противоречащие иерархии лица 1-2 л. > 3 л. [Franks 2017]
- Срб.хрв. ??Toplo =*ti* = ***me*** preporučije, подр. ‘Он(а) тепло тебе меня рекомендовал(а)’
- Болг. *Показват =*му* = ***me***, подр. ‘Они показывают ему тебя’ В древнерусском и среднерусском этот запрет, скорее всего, тоже действовал, но диагностических примеров **вообще нет.**

Отрицательные данные

Какой вывод из отрицательных данных?

- ✓ Нарушение иерархии лица в системе с законом Вакернагеля и дублетными формами местоимений типа *мя ~ меня, относящихся к одной и той же граммеме* (1 л. ед.ч. вин.п.), компенсируется использованием полноударных местоимений.
- ✓ Примеры типа $=ти_2=мя$ ‘тебе меня’, с двумя клитиками, вероятно, могли встречаться в древнерусском и среднерусском, но были низкочастотны.
- ✓ О наличии запрета при отсутствии возможности опросить информантов и отсутствия метаязыковых суждений судить нельзя.

ААЗ и клитики

- В книге 2008 г. ААЗ ввел понятие коэффициента энклитичности — доли употреблений энклитик в дублетных парах местоимений.
- В книге 2007 о «СПИ» ААЗ упоминает, что в чешском переводе Й.Юнгманна (1810 г.)энклитик меньше, чем в русском тексте СПИ. Это наблюдение важно в общем контексте обсуждения гипотезы о том, что гипотетический фальсификатор начала XIX в. мог быть чехом.
- Я попробую посчитать количество сентенциальных клитик в параллельном корпусе СПИ, по данным сайта <http://nevmenandr.net/slovo/>, созданного Борисом Ореховым.

Системы порядка слов с клитиками

- **W-системы.** Стандартные языки с законом Ваккернагеля, базовым порядком X/XP — CL, без требования смежности глагола и клитик.
 - **W⁺-системы.** Языки, где ограничение на постановку клитик от левого конца клаузы накладывается на условие смежности глагола и клитик.
Болгарско-филиппинские W⁺-системы: X/XP — CL — V ~ V — CL.
 - **W*-системы.** Языки, где постановка клитик после начальной составляющей является одной из возможных стратегий.
[Циммерлинг 2013];
- ✓ О правилах рангов славянских языков см. [Franks, King 2000]/

Параллельный корпус

- Выбранные языки: словенский, сербский, хорватский, чешский, словацкий, осетинский (W -системы) болгарский, македонский (W^+ -системы) , польский (W^* -система)
- Во всех этих языках есть цепочки клитик, т.е. объекты, которые строятся по правилу рангов.
- Учитывались разные переводы (при наличии).
- Филологическая точность перевода (наличие ошибок, пропусков, добавлений) не проверялась.
- Не проверялось, стремился ли переводчик к воспроизведению стилистики древнерусского оригинала (экспертная оценка возможна в будущем).

Цепочки и одиночные кластеризуемые клитики

- В соответствии с программой анализа [Zimmerling 2012], опробованной на современном словацком и древнечешском [Циммерлинг 2013: 128 – 153; Zimmerling 2014] языках, отдельно подсчитывалась частотность
 - а) цепочек клитик (две и более клитики, построенные по правилу рангов соответствующего языка).
 - б) одиночные кластеризуемые клитики.

Оригинал, стихи 1-28

- Русский текст СПИ (редакция А.А.Зализняка)
- 11 клауз с сентенциальными клитиками
- 15 кластеризуемых клитик.

Максимум и минимум для стихов 1-28.

- Словенский язык: Плетешерник (1866) – 46 клауз с класт. клитиками, всего 61 класт.клитика; Нахтигаль (1954) – 33/46.
- Польский язык: Красиньский (1856) – 8/9; Белёвский (1906, рифма) – 7/8, Лэпкий (1905) – 10/10, ср. Белёвский (1906, проза) – 14/17, Обрембска-Яблоньска (1954) - 13/17.

Язык	Перевод	год	вид	Кол-во клауз с кластеризуемыми клитиками	кол-во кластеризуемых клитик	Примечания
СЛОВ	Плетешерник	1866	проза	46	61	
СЛОВ	Нахтигаль	1954	проза	33	46	
СРБ	Панић-Суреџ	1857	проза	26	31	
ХРВ	Грчич	1988	проза	27	30	
ХРВ	Бадалич	1972	проза	20	23	
ЧЕШ	Юнгманн	1810	проза	21	24	
ЧЕШ	Кубка	1950	проза	18	22	
ЧЕШ	Врбова	1977	рифма	27	27	
ЧЕШ	Либертин	1978	рифма	18	19	
СЛЦ	Коморовский	1960		15	19	
БЛГ	Стоянов	1954		24	27	
БЛГ	Кадийский	1971		18	23	
МАК	Димитровский	1960		18	21	
ПОЛ	Красиньский	1856		8	9	
ПОЛ	Лэпкий	1905		10	10	
ПОЛ	Белёвский	1906	рифма	7	8	
ПОЛ	Белёвский	1906	проза	14	17	
ПОЛ	Тувим	1950	проза	11	16	
ПОЛ	Обрембска-Яблоньска	1954	проза	13	17	
ПОЛ	Вагилевич	1999	проза	10	13	
ОСЕТ	Плиев	1988	проза	24	27	
ДР.РУС.	Ред. Зализняк	2004	? проза	11	15	

Совершенствование процедуры

- Учет клитик по длине всего памятника (218 стихов)
- Подсчет средних показателей для языка при наличии разных переводов.
- Экспертная оценка переводов с точки зрения их точности и идиоматичности звучания.
- Проверка типичных комбинаций клитик в языках перевода.
- Расширение круга переводов (?)
 - Переведем СПИ на языки варльпири, пашто, тагальский и кашибо-какакаибо?!

Acknowledgements

- При поддержке проектов РФФИ 18-012-00307 «Историческая грамматика русского языка: энциклопедический словарь» и РФФИ 18-512-18003 «Эволюция грамматического строя русского и болгарского языков в сопоставительном аспекте: корпусный анализ и формальные грамматики».
- Благодарю О.И.Беляева за консультацию.

Литература (1)

1. Зализняк 1993 — Зализняк А.А. К изучению языка берестяных грамот // Янин В.Л. , Зализняк А.А. Новгородские грамоты на бересте. Из раскопок 1984 – 1989 гг. М.: Наука, 1993, С. 191 — 319.
2. Зализняк 2004 — Зализняк А. А. Древненовгородский диалект. М., 2004.
3. Зализняк 2007 — Зализняк А.А. «Слово о полку Игореве». Взгляд лингвиста. М., 2007.
4. Зализняк 2008 — Зализняк А.А. Древнерусские энклитики. М., 2008.
5. <http://nevmenandr.net/slovo/>

Литература (2)

6. Ронько 2017 — Ронько Р.В. Ронько Р.В. Надо корова доить! Номинативный объект в севернорусских диалектах // АСТА LINGUISTICA PETROPOLITANA. Труды Института лингвистических исследований РАН / Отв. ред. Н. Н. Казанский. Т. XIII. Ч. 3. — СПб.: Наука, 2017. С. 244-264.
7. Циммерлинг 2013 — Циммерлинг А.В. Системы порядка слов славянских языков в типологическом аспекте. М., 2013.
8. Aissen 2003 — Aissen J. Differential Object Marking: Iconicity vs. Economy // Natural Language and Linguistic Theory 21, 2003. P. 435–483.

Литература (3)

9. Franks, King 2000 — Franks S., King T.L. A Handbook of Slavic Clitics. Oxford: OUP, 2000.
10. Franks 2017 — Franks S. Syntax and Spell-Out in Slavic. Bloomington, Indiana: Slavica, 2017.
11. Zimmerling 2012 — Zimmerling A. A unified analysis of clitic clusters in the world's languages // Computational linguistics and intellectual technologies, Vol. 11 (18). P. 726 — 738.
12. Zimmerling 2014 — Zimmerling A. Clitic Templates and Discourse Marker TI in Old Czech // J.Witkoś, S.Jaworski (eds.). New Insights in Slavic Linguistics. Frankfurt-am-Main: Peter Lang. P. 393 — 406.