

Некоторые применения математических
методов в языкознании, МГУ,
18 октября 2014

Валентностные классы двухместных глаголов в языках Европы (и не только)

Квантитативно-типологическое
исследование

Сергей Сай
ИЛИ РАН, Санкт-Петербург
serjzhka@yahoo.com

Введение

- Продолжающийся проект!
- В большинстве языков можно идентифицировать **базовую переходную конструкцию**
 - Это рабочее предположение, связано с определенными допущениями, далее обсуждаться здесь не будет

Введение

- Кодирование участников А и Р при этом лингвоспецифично (нет универсалий)

	А	Р
Английский	В начале клаузы, маргинальное согласование («indexing»), именительный падеж у местоимений («flagging»)	Позиция за глаголом, косвенный падеж у местоимений («flagging»)
Монгольский	Именительный падеж («flagging»)	Винительный падеж («flagging»)
Абхазский	Эргативное согласование («indexing»)	Абсолютное согласование («indexing»)

Введение

- Семантическая база «высокой переходности» хорошо изучена: (Hopper & Thompson 1980 и далее Tsunoda 1981; Dixon, Aikhenvald 2000; Kittilä 2002; Malchukov 2006; Næss 2007 ...)
- **Грамматические** факторы отклонения от переходности:
 - пассив и другие залогов,
 - хаби́туалис,
 - партитивность,
 - отрицание,
 - некоторые зависимые клаузы, и т.д.

Введение

- **Лексические** факторы отклонения от переходности: основной объект изучения здесь
- Часто упоминаются в «отрицательных» контекстах: глаголы не являются переходными, т.к.:
 - «When [the effectiveness condition] is not met, ERG-ABS may (though not always) fail to occur, and we will have **some other case frames**» (Tsunoda 1981: 393)
 - «The selection principles apparently only govern argument selection for two-place predicates having a subject and a true direct object» (Dowty 1991: 576).
- Двухместные непереходные глаголы распадаются на валентностные классы (меньшие, чем переходный класс)

Цели и задачи

- Насколько стабильны кросс-лингвистически многоместные валентностные классы?
- Насколько универсально распределение по валентностным классам определяется семантически?
- Существуют ли «естественные» валентностные классы (ср. «естественные реципроки»)?
- Насколько в изучаемом отношении похожи друг на друга разные языки?
- Как все это измерить?

Теоретический контекст

- Есть типологические исследования по
 - отдельным косвенным позициям, например о «дативе» (van Belle & van Langendonck 1996),
 - стратегиям грамматикализации, e.g. От пространственных значений к ядерным актантам (Ганенков)
 - семантическим классам предикатов, e.g. экспериенциальным (Bossong 1998).
- Подробные исследования систем валентностных классов в отдельных языках: Levin 1993 (английский) и Апресян 1967 (русский)
 - Во многом это источник вдохновения

Другие похожие исследования

- Leipzig Valency Classes Project
- Продолжающиеся исследования Бальтазара Бикеля и коллег
 - Между этими проектами и нашим исследованием есть существенные различия в подходах и методах, сегодня они за рамками рассмотрения

Техническая сторона проекта

- Анкета создана на основе пилотных исследований в соответствии с такими ожиданиями:
 - предикаты будут как минимум 2-местными (X и Y);
 - будет входить много глаголов, которые типологически склонны к непереходности
- На данном этапе игнорируется вариативность кодирования актантов
 - В инструкции говорится о том, как действовать, если получено более одной модели

Техническая сторона проекта

- Для каждого предиката в каждом языке фиксируются морфосинтаксические средства кодирования 2 участников (X и Y)
- Глаголы признаются принадлежащими к одному и тому классу тогда и только тогда, когда при них совпадает кодирование X-ов и Y-ов
- Иногда данные не удается получить (пробелы):
 - Например, невозможно естественно выразить искомое значение;
 - X или Y не оказывается зависимым на уровне клаузы,
 - и т.д.

Какие языки изучены

- На данный момент собраны данные для
 - 34 языка Европы (в том смысле, как это понималось в проекте EUROTYP)
 - 9 других языков

Какие языки изучены

баскский	[Н. Заика]	сербский	[А. Макарова]
испанский	[Е. В. Горбова]	польский	[Г. Мороз]
французский	[Е. Е. Корди]	украинский	[Н. Заика]
итальянский	[А. Александрова]	русский	[С. Сай]
румынский	[Д. Суетина]	ингерманландский	[Д. Мищенко]
ирландский	[Д. Николаев]	эстонский	[И. Кюльмоя]
норвежский (букмол)	[О. Кузнецова]	коми зырянский	[Е. Сергеева]
английский	[Д. Николаев]	коми пермяцкий	[Е. Сергеева]
нидерландский	[М. Князев]	эрзянский (морд.)	[К. Шагал]
немецкий	[С. Бирцер]	башкирский	[С. Сай]
литовский	[Н. Заика]	азербайджанский	[Л. Курбанова]
латышский	[Н. Перкова]	калмыцкий(*)	[С. Сай]
албанский	[В. Дивеева]	осетинский	[А. Выдрин]
древнегреческий*	[И.Ибрагимов]	цахурский*	[Д. Герасимов]
новогреческий	[Е. Желтова]	багвалинский*	[Д. Герасимов]
армянский	[В. Крылова]	ингушский	[Дж. Николз]
цыганский (кэлдэрарский)	[К. Кожанов]	лезгинский	[Р. Мамедшахов]

*Не было полевой работы. Данные взяты из опубликованных источников.

Какие языки изучены

арабский	[Р. Мамедшахов]	японский	[Ю. Конума]
лоома	[Д. Мищенко]	кхмерский	[С. Дмитренко]
гуро	[О. Кузнецова]	китайский	[Е. Колпачкова]
тувинский	[С. Оскольская]	гуарани	[Д. Герасимов]
чукотский	[М. Пупынина]		

- Проект является коллективным: в полученных выводах есть вклад всех названных лиц
- Что общего с традиционным подходом Ленинградской типологической школы?
 - Единая анкета, единый подход, исследователи занимаются языками, в которых они разбираются
 - Но нет полноценной процедуры создания выборки (привлекаем всех, кого можем привлечь)

Два ВОЗМОЖНЫХ взгляда

- Сравнение предикатов (часть 1)
- Сравнение языков (часть 2)

Сравнимая предикаты: переходность

- Коэффициент переходности глагольного значения = число языков, где значение выражены переходной структурой, поделенное на общее число языков, для которых были получены релевантный данные

Например 'целовать':

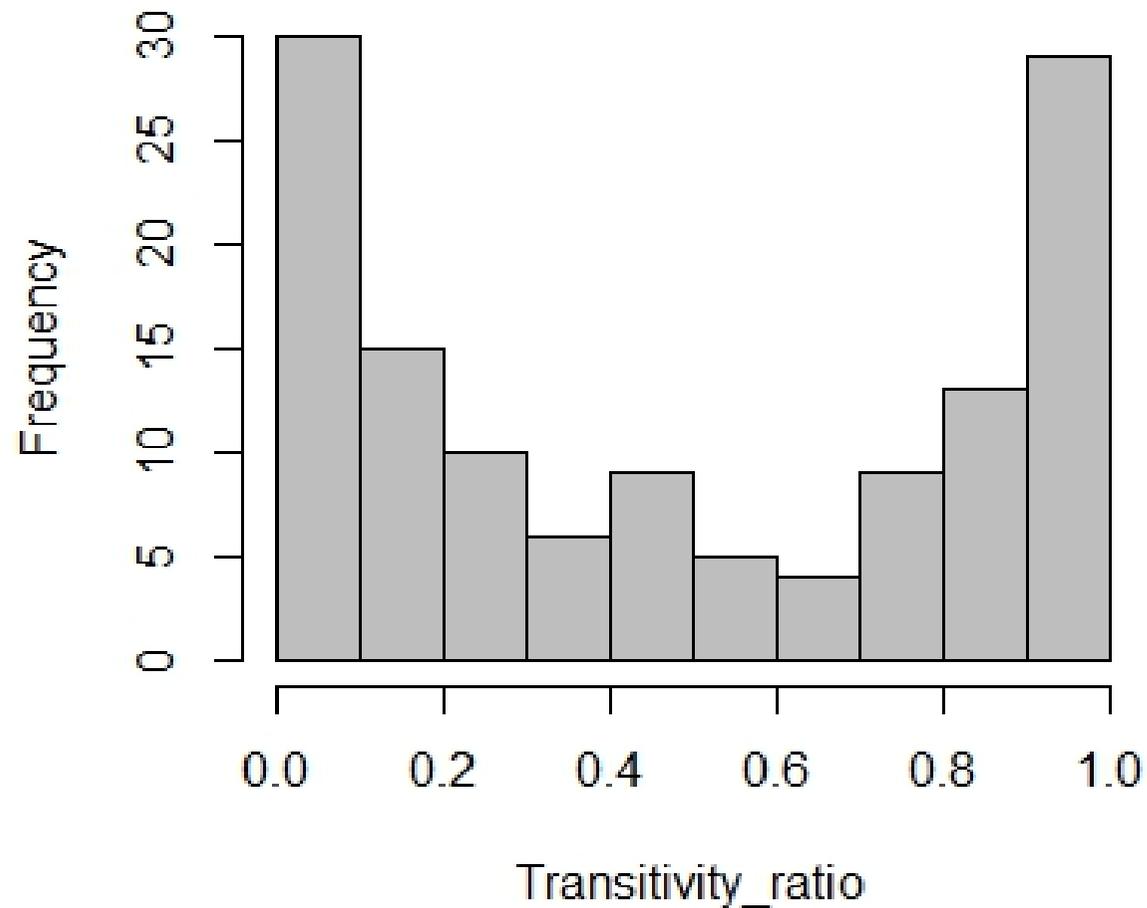
- выражается переходной структурой в 31 языке выборки
- непереходной структурой в 10 языках выборки
- данных не хватает для 2 языков

$$\text{TransitivityRatio ('целовать')} = 31 / (31 + 10) = 0.76$$

have enough	0	be content	0.07	meet with	0.29	cost	0.70	milk	0.94
fight	0	have a ache	0.08	influence	0.29	catch up	0.71	paint	0.94
be different	0	smell	0.09	obey	0.33	love	0.72	pour	0.97
fall behind	0	have P left	0.10	sympathise	0.36	hate	0.73	open	0.97
cut oneself	0	enter	0.11	enjoy	0.36	remember	0.74	take off	0.97
get stuck	0	tell (sb)	0.11	envy	0.37	kiss	0.75	drive	0.97
speak with	0	be marvelled	0.12	help	0.39	move (finger)	0.78	throw	1
get mixed	0	be afraid	0.14	reach	0.40	hit	0.81	take	1
have a quarrel	0	be called	0.14	miss	0.40	despise	0.81	bend	1
drown	0	think	0.14	dream	0.40	want	0.81	eat	1
be angry	0	be similar	0.15	forget	0.42	look for	0.82	fry	1
be cross	0	lack	0.17	flatter	0.45	respect	0.82	make	1
become upset	0	fall in love	0.17	beat (in a game)	0.45	punish	0.82	try to catch	1
get irritated	0	trust	0.17	attack	0.46	bite	0.84	break	1
agree	0.03	like	0.18	taunt	0.47	leave	0.86	wash	1
dismount	0.03	shoot at	0.18	wave	0.47	hear	0.86	put on	1
depend	0.03	be squeamish	0.18	follow	0.48	call	0.86	plough	1
rejoice	0.03	believe	0.19	touch	0.50	give birth	0.86	sing	1
be ashamed	0.03	have (illness)	0.20	avoid	0.50	see	0.87	write	1
be friends	0.04	dream	0.21	play (guitar)	0.52	surround	0.87	drink	1
take offence	0.04	hit (target)	0.23	forfeit	0.52	cross	0.89	melt (tr.)	1
fill (intr.) (with)	0.06	answer	0.24	have got	0.54	understand	0.89	cover	1
go out	0.06	get to know	0.25	love (inan.)	0.59	know	0.89	lose	1
be surprised	0.06	look	0.27	manage	0.67	drop	0.91	kill	1
go well	0.07	need	0.27	listen	0.69	find	0.92	read	1
lose to	0.07	be fond	0.27	wait	0.70	hold	0.92	upset	1

Сравнивая предикаты: переходность

Transitivity ratios of verbs (37 lgs)



Сравнимая предикаты: переходность

- Отдельные предикаты обычно отчетливо предпочитают переходность или непереходность.
 - Внимание: это, разумеется, всецело зависит от (произвольного) выбора предикатов!
 - Но с самого начала мы старались включить как можно больше предикатов, про которые было известно, что они могут оказаться где-то на середине шкалы.

Сравнимая предикаты: переходность

- Идеи предшественников: имплицитивные иерархии, ср:

«1a) direct effect (*kill, break* subtype) > 1b) direct effect (*hit, shoot* subtype) > 2a) perception (*see* subtype) > 2b) perception (*look* subtype) > 3) pursuit (*search, wait*) > 4) knowledge (*know, understand, remember, forget*) > 5) feeling (*love, like, want, need*) > 6) possession (*have*)» [Tsunoda 1981].

- Ср. с нашими данными:

1a) direct effect	break	1
1b) direct effect	hit	0.82
	shoot	0.19
2a) perception	see	0.86
	hear	0.86
2b) perception	look	0.28
	listen	0.71
3) pursuit	search	0.84
	wait	0.69
4) knowledge	know	0.89
	understand	0.88
	remember	0.70
	forget	0.4
5) feeling	love	0.71
	like	0.16
	need	0.22
6) possession	have	0.56

Сравнимая предикаты: переходность

- Не считая глаголов группы «direct effect» иерархия Цуноды довольно плохо отражает данные.
- Более того, внутри выделенных семантических групп в основном нет большого единства.
 - Дисклеймер: наши данные очень нерепрезентативны географически, но все же...

Сравнимая предикаты: профили переходности

- Для каждой пары предикатов можно установить расстояние Хэмминга.

Е.г. для 'достичь' и 'дотронуться' (столбцы соответствуют языкам; "t" -- переходный, "i" – двухместный непереходный, "?" – нет данных):

[21] 'reach' itti?ii**itttiit**tit?ti?i...

[22] 'touch' itttiiii**tttiit**tttiit...

- На основании этих 20 языков, нормализованное расстояние между 'достигать' и 'дотронуться' равно **0.29** (=5/17).

Профили переходности

- Матрица расстояний
- NeighborNet: двухмерная визуализация разветвлений таксонов (в нашем случае предикатов).
- SplitsTree4 software (Huson, Bryant 2006): изначально применялся для филогенетических задач

NeighborNet: 130 предикатов, профили переходности в 38 языках

Более переходные глаголы



drop remember
want@object@
leave look_for
read, throw, cover listen
give_birth
write despise
open bite call
surround hate
sing
paint drink
milk hold wait
take_off move@bodypart@
drive eat cost manage
put_on know respect
make hear kiss
break see catch_up
fry kill love@inanimate@
plough cross love
washpour
understand
try_to_catch
upset find
melt@tr@
take
bend

meet_with
be_fond
sympathise help
obeyenjoy look believebe_surprised
forget taunt influence think be_friendsget_irritated
beat wave trust go_wellbear_a_grudge
play@instrument@ dream@sleeping@ envy
miss
lack be_marvelled
have_enough
hit@target@ cut_oneself, get_stuck
need answerfill@intr@_with
have@illness@ have__left
be_afraid shoot_at
get_to_know enter fight
fall_in_love
like smell
be_called
be_similar
tell

have_a_quarrel
dismount, take_offence
fall_behind
be_content
be_ashamed
speak_with
have_a__ache
lose_to get_mixed
be_squeamish rejoicebecome_upset
agree be_angry
dream
obeyenjoy look believebe_surprised
think be_friendsget_irritated
trust go_wellbear_a_grudge
depend
go_out
be_marvelled
hit@target@ cut_oneself, get_stuck
need answerfill@intr@_with
have@illness@ have__left
be_afraid shoot_at
get_to_know enter fight
fall_in_love
like smell
be_called
be_similar
tell



Менее переходные глаголы

Профили переходности

- Дендрограмма получилась почти одномерной, это измерение в целом соответствует степени склонности к переходности.
- Эти данные косвенно подтверждают идею существования иерархии переходности!

Локус непереходности

- Какой (какие) из актантов А, Р попадает в неядерную позицию?

ЛИТОВСКИЙ:

А-локус:

(1) *Petr-ui* *patink-a* *šit-ie* *marškini-ai*
Пятрас-DAT.SG нравится-PRS.3 этот-NOM.PL рубашка-NOM.PL
'Пятрасу (А) нравится его рубашка (Р).'

Р-локус:

(2) *Petr-as* *atsilik-o* *nuo* *Marij-os*
Р.-NOM.SG fall.behind-PST.3 from М.-GEN.SG
'Пятрас (А) отстал от Марии (Р)'

А&Р-локус (двойной локус):

(3) *Petr-ui* *pakank-a* *pinig-ų*
Пятрас.-DAT.SG хватать-PRS.3 деньги-GEN.PL
'Пятрасу хватает денег.'

Локус непереходности

	No.	%		No.	%
transitive	1658	49,8%	P-locus	1469	44,1%
A-locus	167	5,0%	A&P-locus	38	1,1%

- Двойной локус (A&P-локус) наблюдается редко (1.1%).
Объяснение Мальчукова (2006): “Primary argument immunity principle”
- Он встречается даже реже, чем ожидалось бы при гипотезе о независимости A-локуса и P-локуса (хи-квадрат, $p < .001$)
- В наших данных структуры с A-локусом фиксируются значительно реже структур с P-локусом

Локус непереходности

- Глаголы с высокой долей А или А&Р-локуса:

have enough P	92,6%	be called	17,2%
have a P-ache	88,5%	forget	14,3%
have P left	80,6%	love	13,9%
lack	80,0%	see	10,5%
like	78,8%	hate	10,0%
have	42,9%	be surprised	10,0%
need	42,4%	be ashamed	10,0%
dream (sleeping)	28,0%	smell	9,1%
have (illness)	24,0%	be fond	9,1%
sympathise	21,4%	find	8,3%

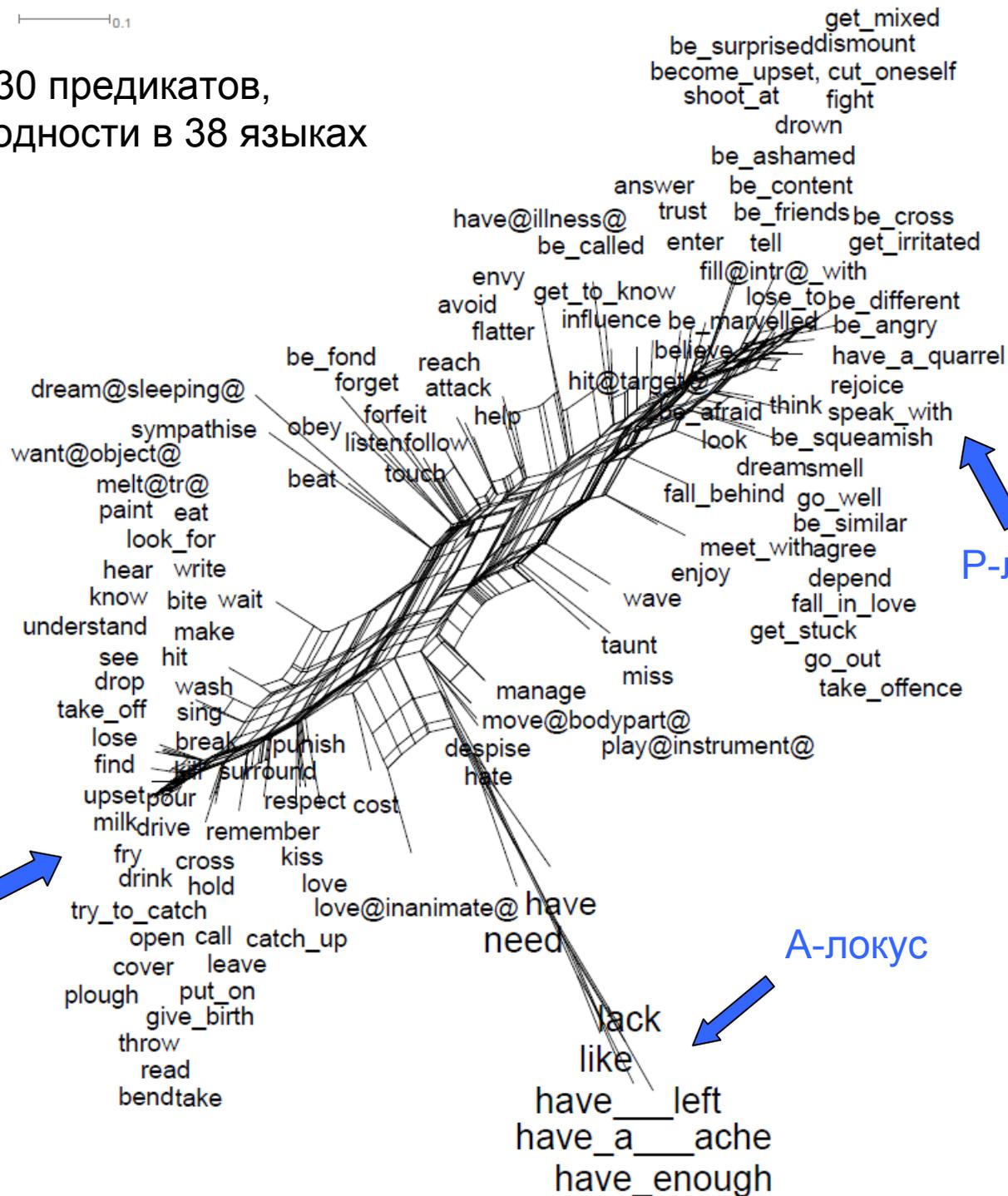
Локус непереходности

- Склонность к А или А&Р-локусу:

Предикаты, связанные с обладанием >
(некоторые) экспериенциальные предикаты >
другие предикаты

0.1

NeighborNet: 130 предикатов, локус непереходности в 38 языках



Переходные

P-локус

A-локус

Сравнивая предикаты: классы

- Для каждой пары предикатов устанавливается расстояние Хэмминга: количество таких языков, в которых два предиката попадают в разные классы:

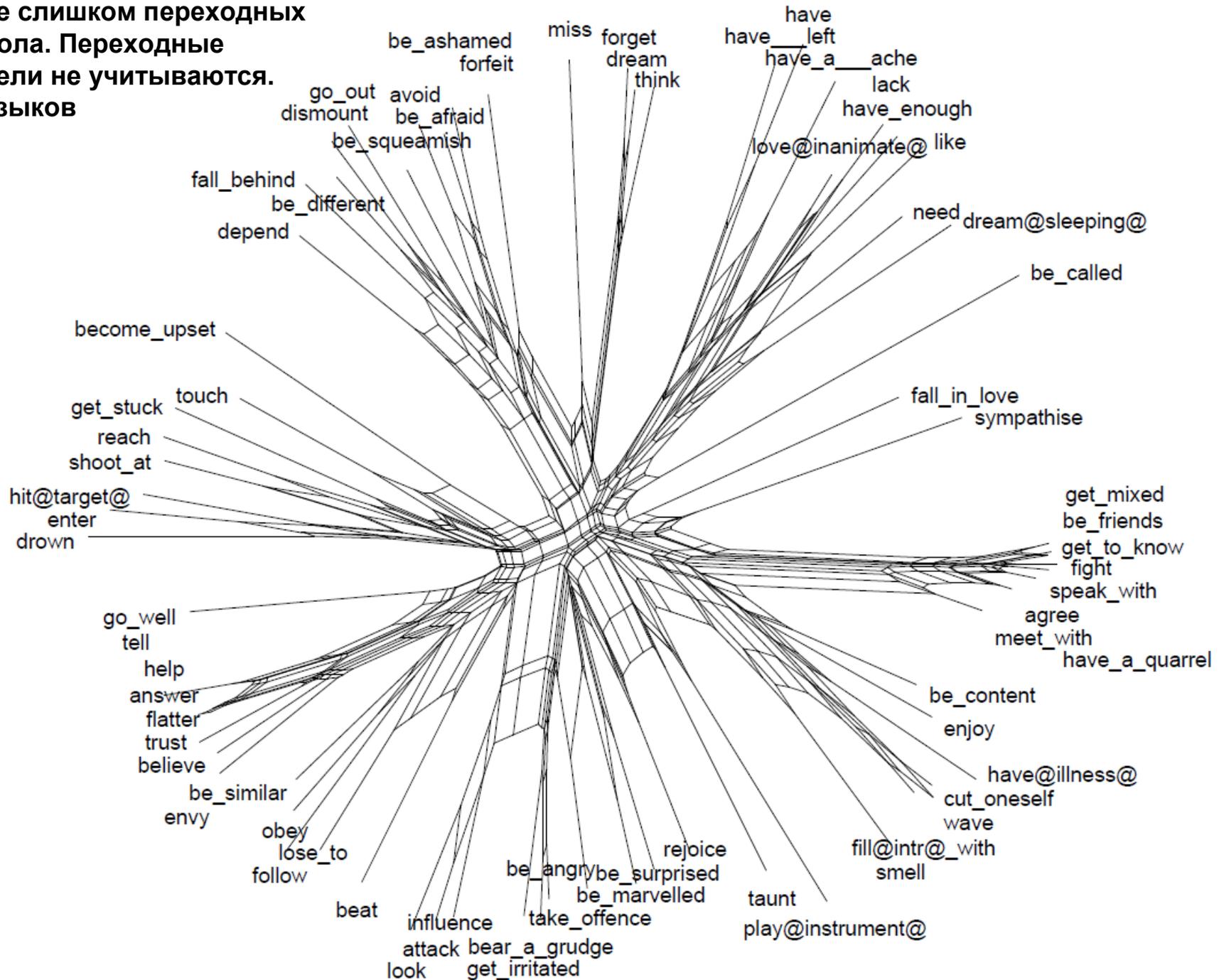
`бояться' r l a g m a l f a t d g j
`избегать' r t ? g ? a n f t ? d g f... etc.

- Буквы – это шифры для отдельных валентностных классов в отдельных языках (т.е. столбцах); “?” = нет данных

Сравнивая предикаты: только непереходные структуры

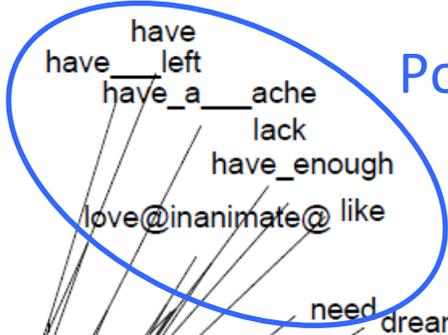
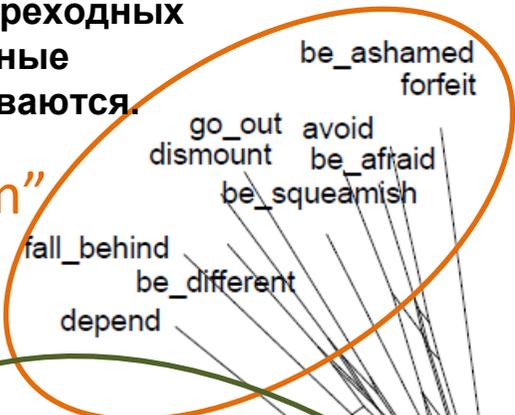
- Отбросим значения, которые в 50% языков или более выражаются переходными структурами.
- Считаем глаголы принадлежащими к одному и тому же классу в определенном языке если это был какой-либо непереходный класс.
- Результаты:

74 не слишком переходных глагола. Переходные модели не учитываются. 37 языков

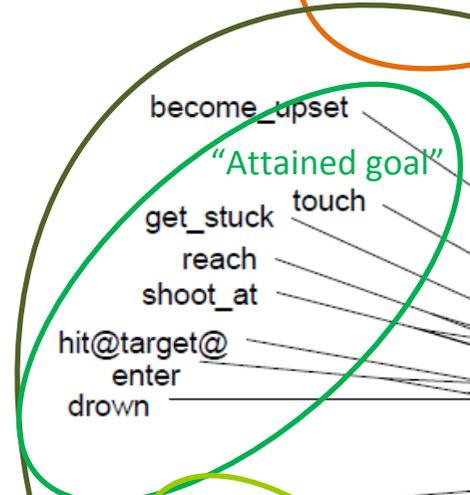


74 не слишком переходных глагола. Переходные модели не учитываются. 37 языков

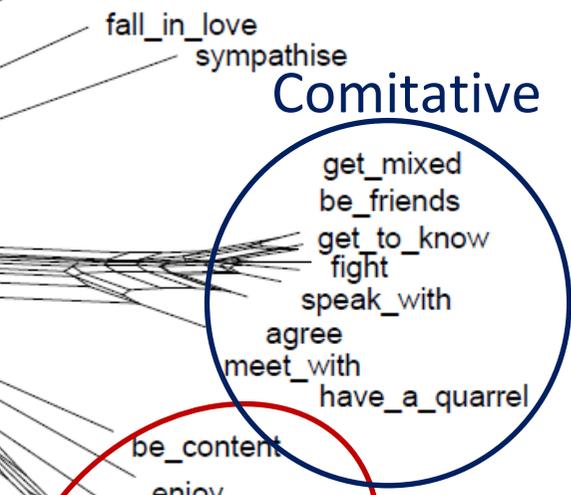
“Motion from”



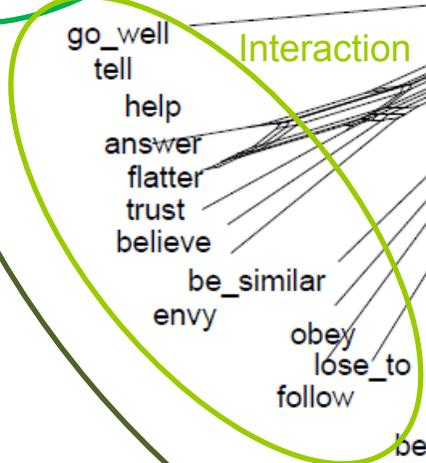
Possession



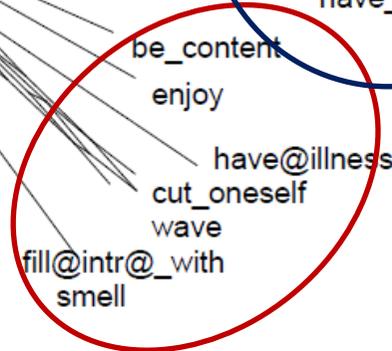
“Attained goal”



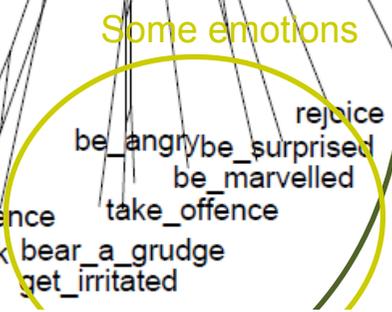
Comitative



Interaction



Instrument/Cause



Some emotions

“Motion towards”

Сравнивая предикаты

- Очевидно, существуют кластеры значений, которые кросс-лингвистически часто попадают в один и тот же валентностный класс
- Можно ли говорить, что это эмпирическое основание для абстрактных семантических ролей (“Meso-roles”)?
- См. (Bickel et al. 2011 “Semantic role clustering: an empirical assessments of semantic role types”) и (в печати) – более сложная математика, та же проблема
- “Meso-roles” – это такие обобщенные роли, которые более абстрактны, чем роли при отдельных предикатах, но менее абстрактны, чем такие обобщенные роли, как S, A, P, T, G

Сравнивая предикаты: Экспериенцер как «мезо-роль» (Meso-role)

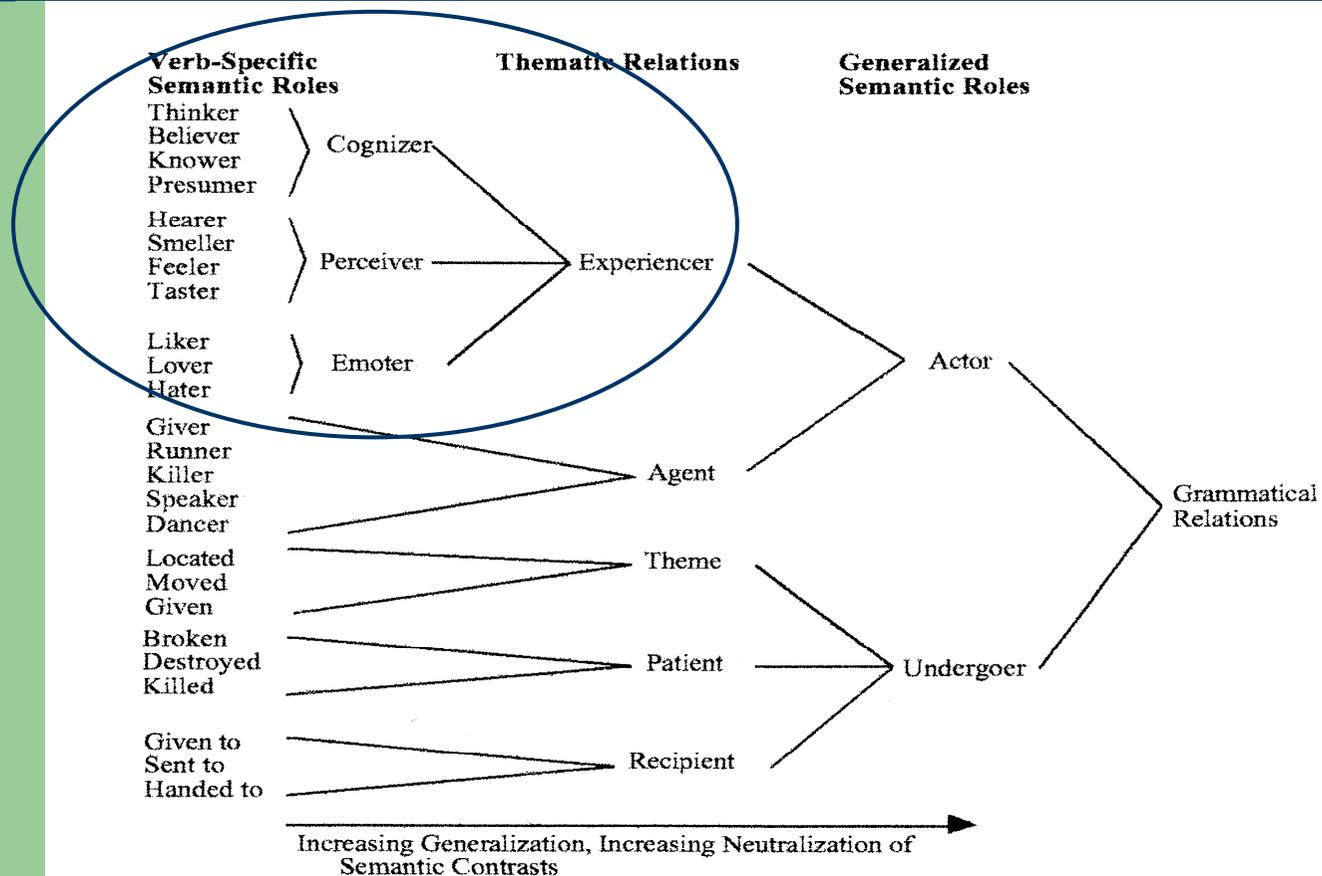
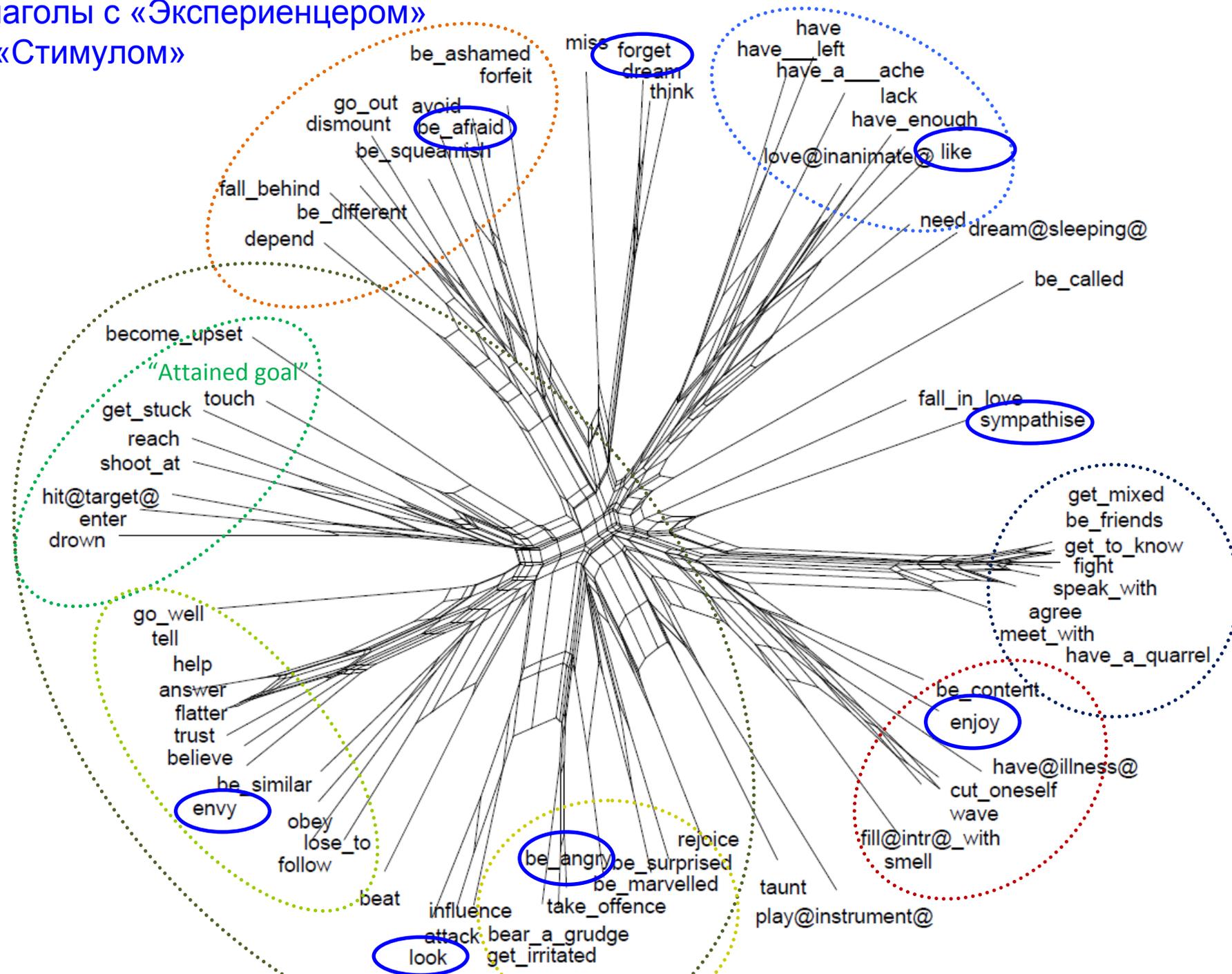


Figure 1: Relation of generalized semantic roles to thematic relations

Сравнивая предикаты: Экспериенцер как «мезо-роль» (Meso-role)

- Полученные нами кандидаты на «роли» не всегда совпадают с привычными «семантическими ролями».
- См. расположение традиционно понимаемых глаголов с Экспериенцерами в полученном NeighborNet.

Глаголы с «Экспериенцером» и «Стимулом»



Предсказуемость моделей

- Попадание глаголов в определенные классы в конкретных языках сочетает элементы межъязыковой регулярности и конкретно-языковой случайности.
- В одинаковой ли степени регулярно / случайно кодирование разных предикатов?

Предсказуемость моделей

- Для того чтобы измерить степень закономерности попадания предиката в тот или иной класс необходимо *tertium comparationis*. Что подойдет?
 - «Семантические роли», определенные на априорных основаниях, не подходят. Что тогда?
 - Валентностные свойства в других языках!

Предсказуемость моделей

- Допустим есть только 4 глагола, попадающих в определенный класс в отдельном языке L1:

	L1
V_i	A
V_j	A
V_k	A
V_l	A
...	

Предсказуемость моделей

- Как насчет их модели управления в L2?

	L1	L2
V_i	A	W
V_j	A	W
V_k	A	W
V_l	A	W
...		

- Модель управления этих глаголов в L2 выглядит очень предсказуемой!

Предсказуемость моделей

- Как насчет их модели управления в L2?

	L1	L2
V_i	A	X
V_j	A	Y
V_k	A	Z
V_l	A	W
...		

- Модели управления этих глаголов в L2 кажутся совершенно непредсказуемыми!

Предсказуемость моделей

- Реальный пример

	Russian	Kalmyk
'be afraid'	NOM_GEN	NOM_ABL
'reach'	NOM_GEN	NOM_DAT
'avoid'	NOM_GEN	NOM_ACC
'forfeit'	NOM_GEN	NOM_ABL
'be ashamed'	NOM_GEN	NOM_ABL

- Исходя из перспективы данных русского языка, валентностный класс калмыцких глаголов 'reach' и 'avoid' кажется менее закономерным, чем для трех других русских глаголов с генитивным объектом

Предсказуемость моделей: формализация

- Отдельный предикат, два данных языка:
условная вероятность:

$$\pi(V_i)(L_j \rightarrow L_k) = p(\text{Class}(V_i, L_k) | \text{Class}(V_i, L_j))$$

e.g.

$$\pi(\text{'reach'}) (\text{Russian} \rightarrow \text{Kalmyk}) = 1/5 = 0.2$$

$$\pi(\text{'be_afraid'}) (\text{Russian} \rightarrow \text{Kalmyk}) = 3/5 = 0.6$$

Предсказуемость моделей: формализация

- Конкретный предикат, один данный язык: нужно рассмотреть столько “перспектив” (языков выборки), сколько возможно:

$$\pi(V_i)(L_k) = \frac{\sum_{j=1}^n \pi(V_i)(L_j \rightarrow L_k)}{n}$$

e.g. π ('be_afraid') (Kalmyk) = 0.54

Предсказуемость моделей: формализация

- Конкретный предикат, разные языки:
среднее значение

$$\pi(V_i) = \frac{\sum_{j=1}^n \pi(V_i)(L_j)}{n}$$

e.g. π ('be_afraid') = 0.42

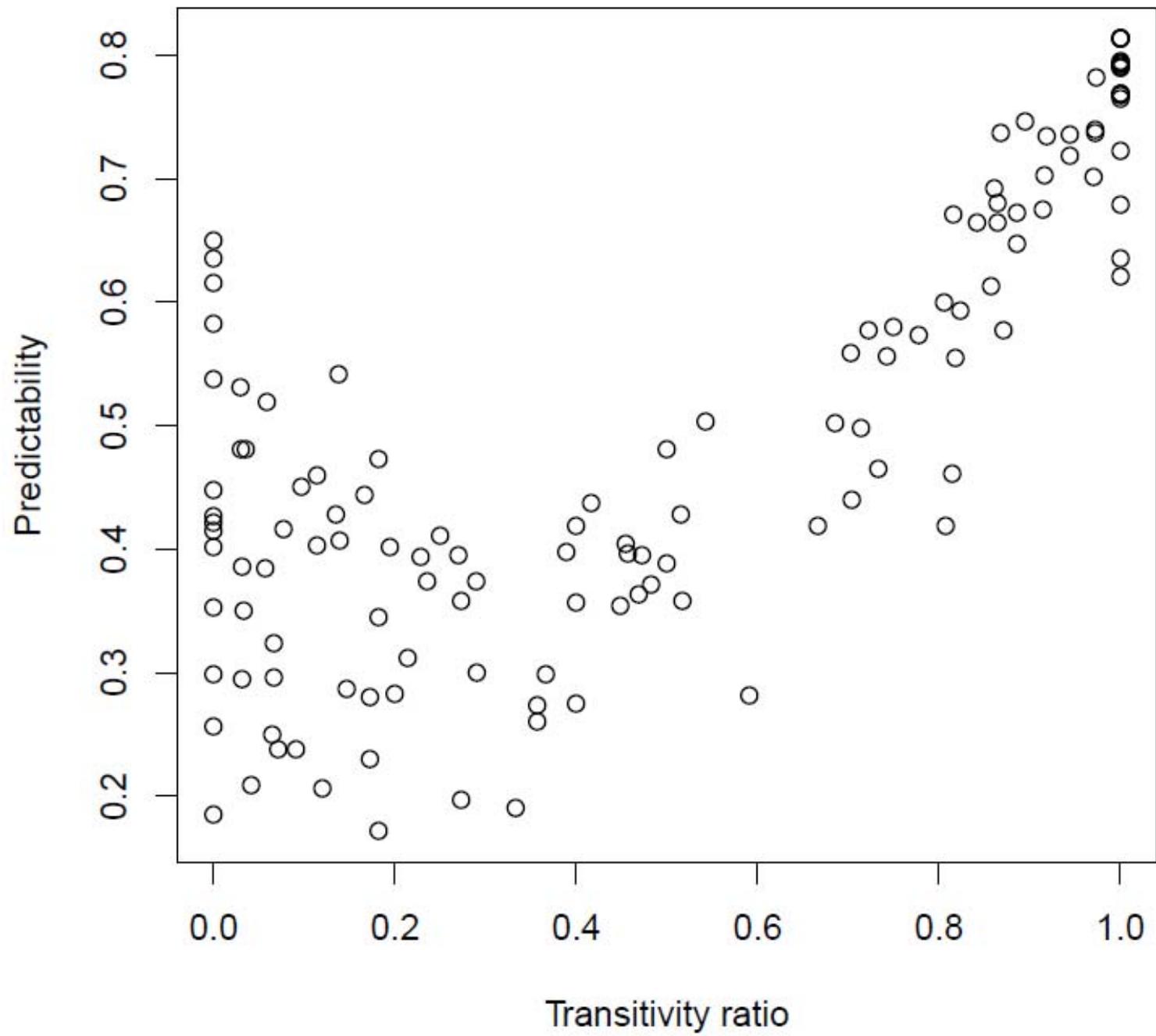
Предсказуемость моделей: формализация

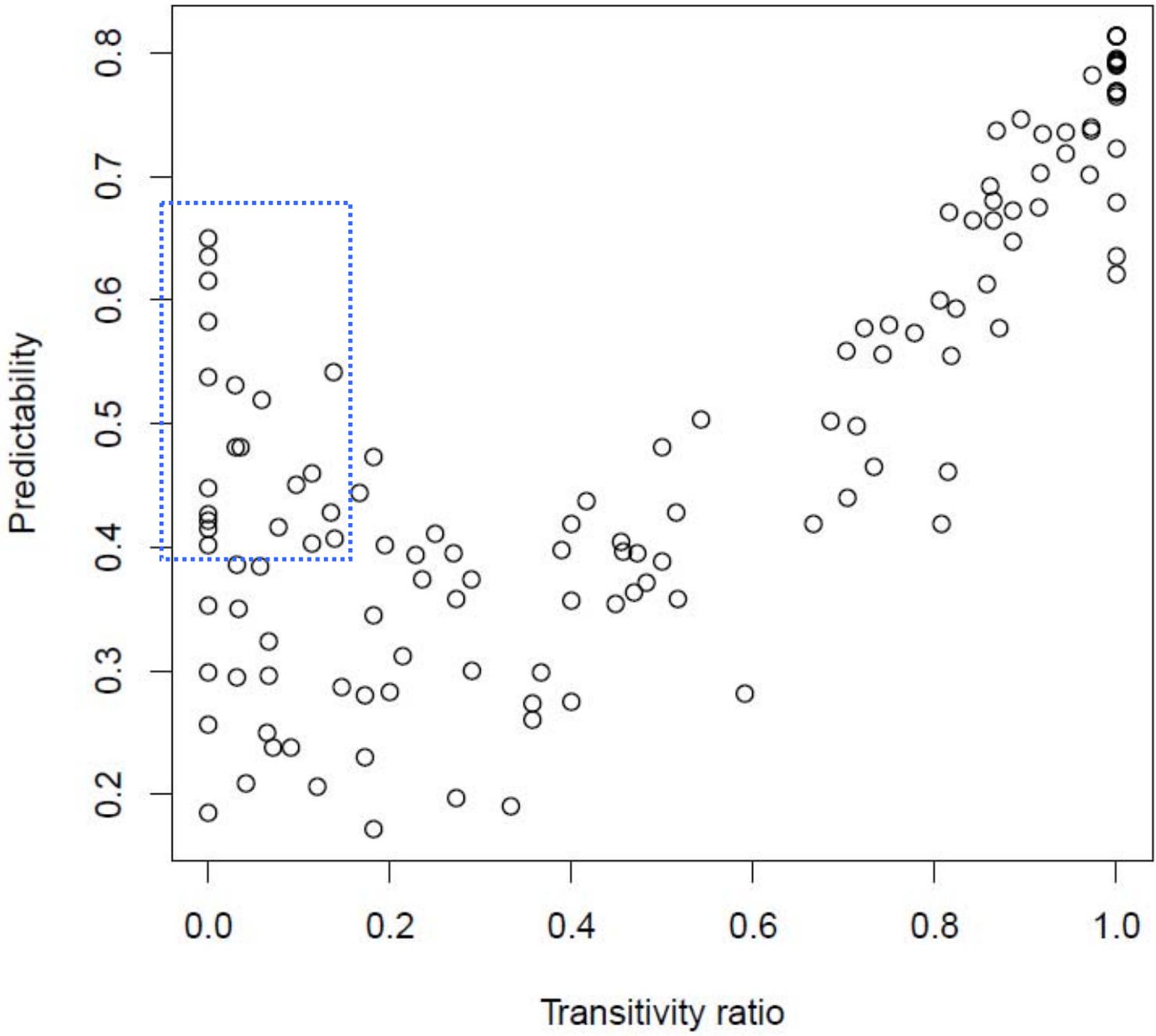
- Идентификация того, какой именно класс мы изначально считаем переходным, для такой техники нерелевантна!
- Сравнение с (fuzzy) cluster analysis
 - Минус: в результатах не видны собственно кластеры.
 - Плюс: мы не исходим из презумпции, что кластеры **должны быть**. Для каждого предиката мы определяем, насколько у него **стабильное** «соседство» по валентностным классам, а не то, в какой степени он принадлежит конкретному кластеру.

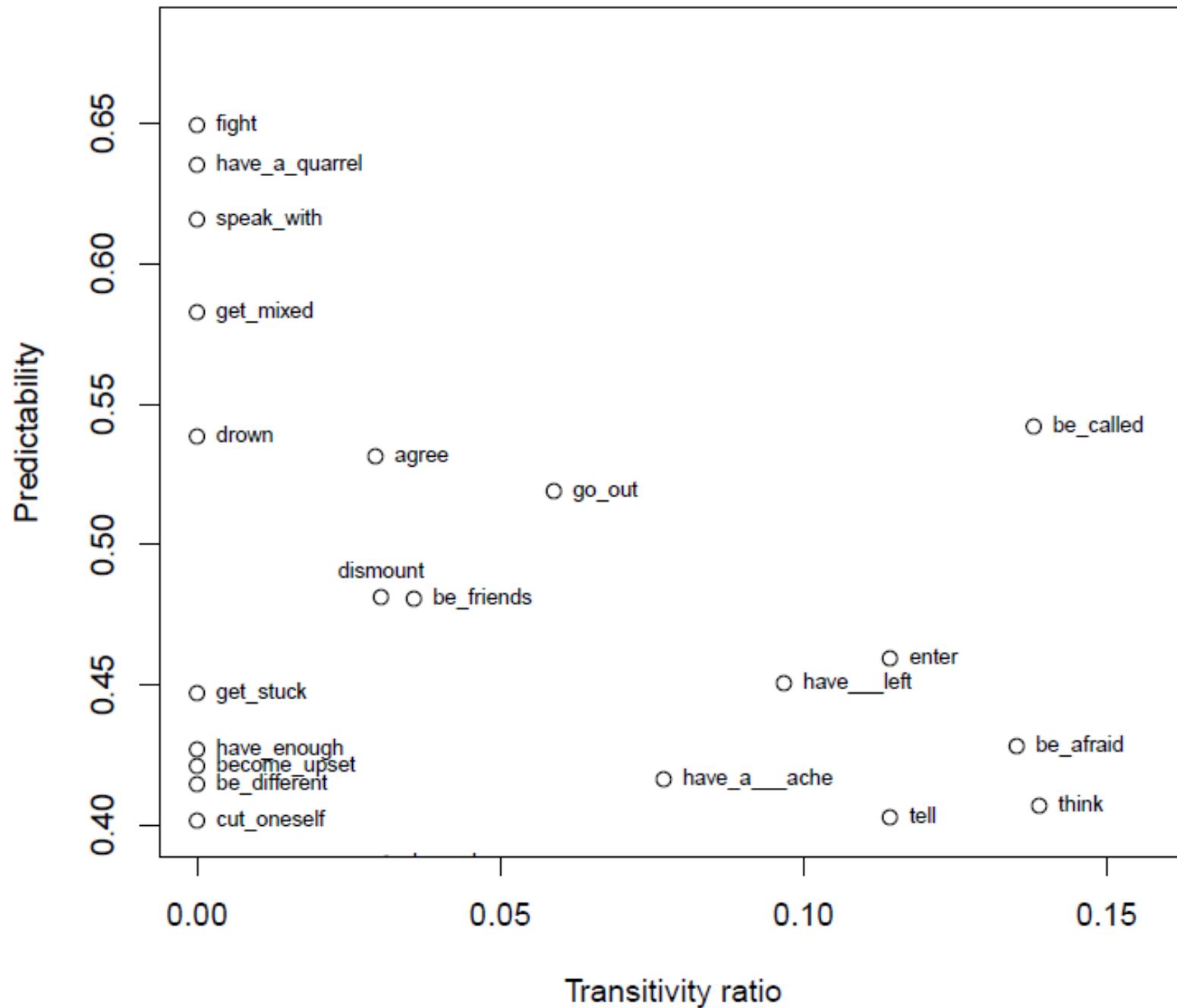
Предсказуемость моделей: результаты

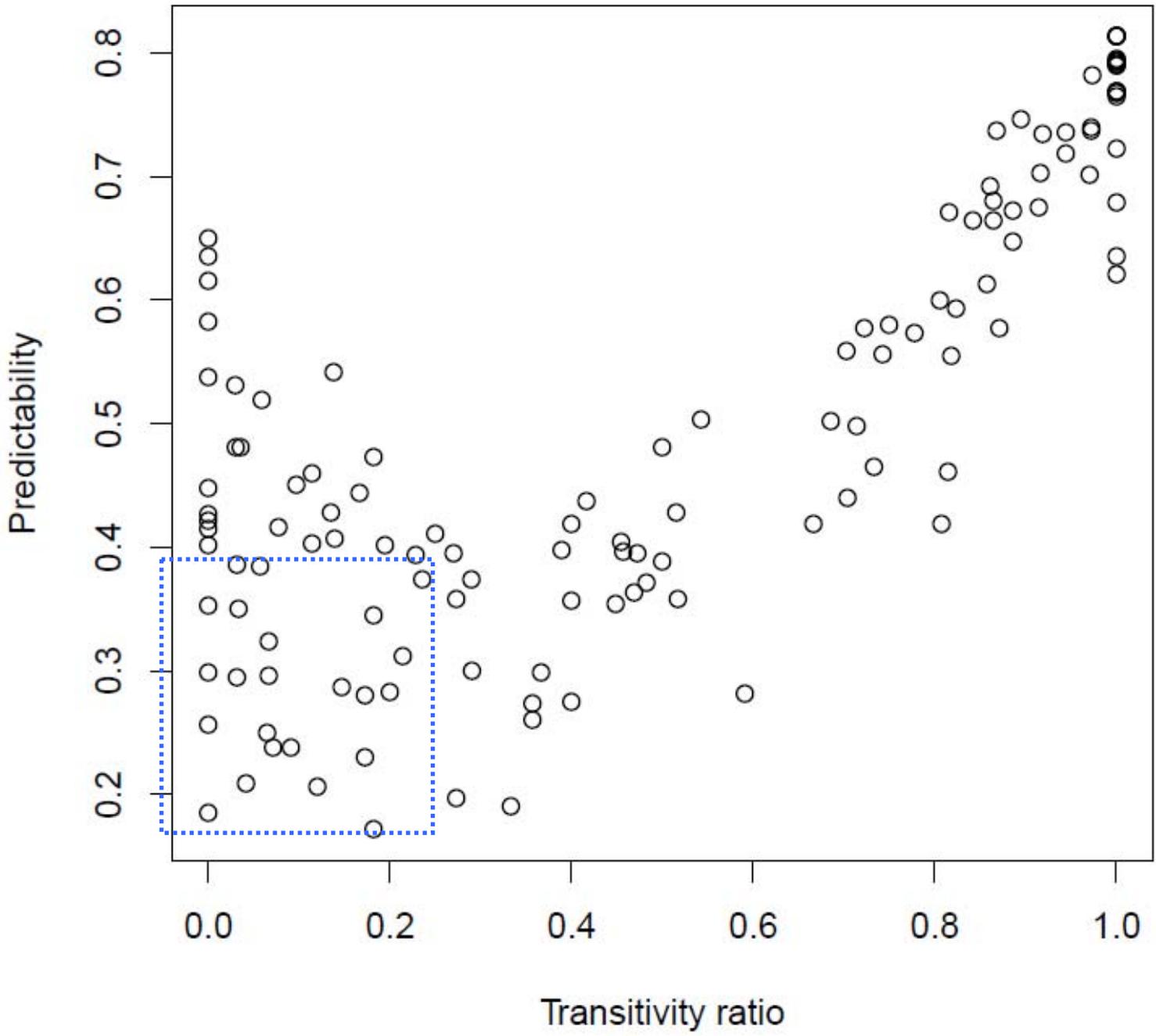
- Выполнено в R Машей Овсянниковой
- Как и ожидалось, глаголы с высоким коэффициентом переходности обнаруживают высокую предсказуемость.
- Предикаты с $\pi > 0.75$:

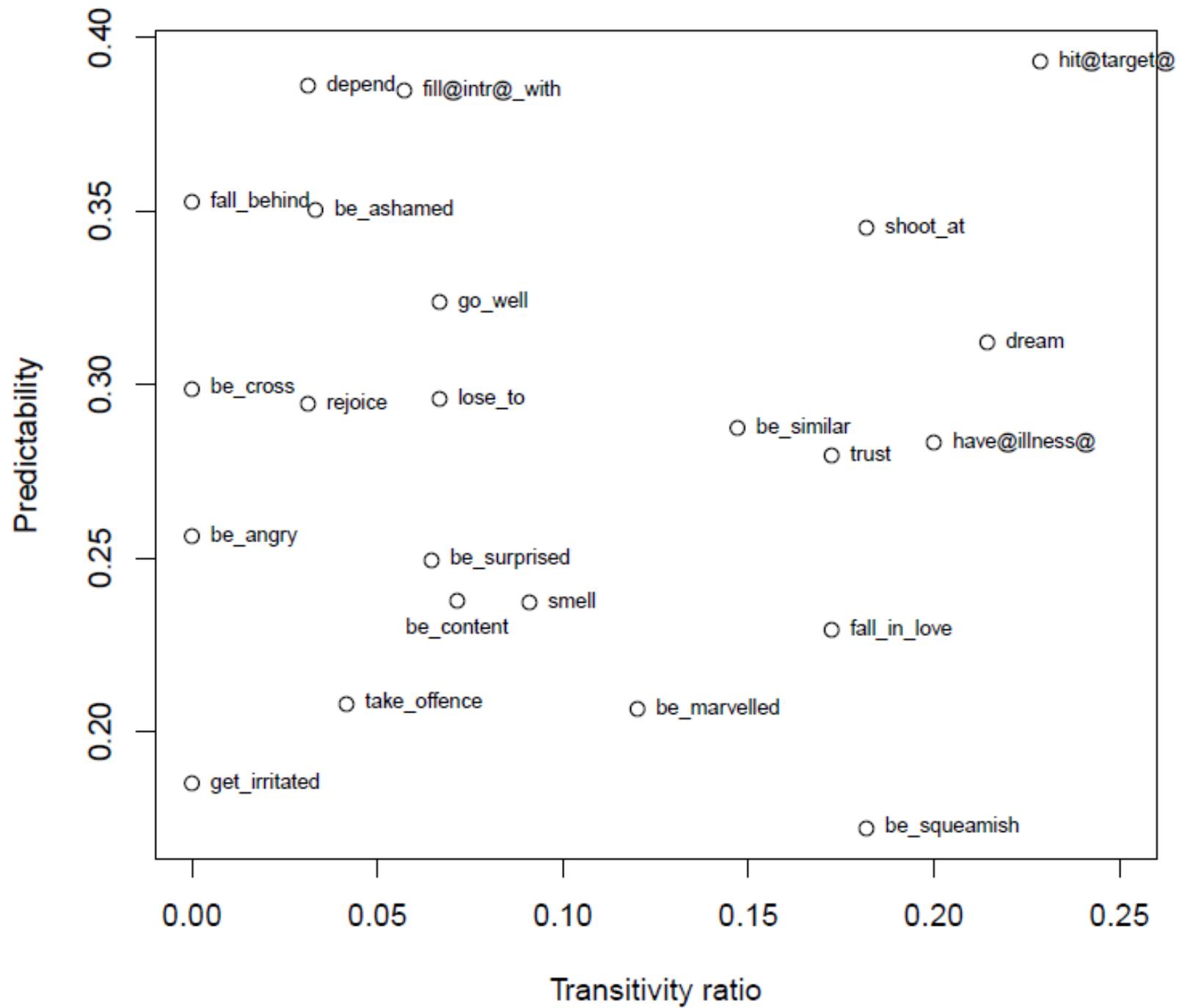
'break'	0.81	'read'	0.79	'wash'	0.77
'make'	0.81	'eat'	0.79	'sing'	0.77
'take'	0.81	'drink'	0.79	'plough'	0.77
'write'	0.81	'bend'	0.79	'lose'	0.77
'throw'	0.80	'kill'	0.79	'fry'	0.77
'put on'	0.79	'drive'	0.78		











Предсказуемость моделей

- Низкая переходность, высокая предсказуемость:
 - Симметричные предикаты ('драться', 'разговаривать с', 'смешаться', 'согласиться')
 - Некоторые глаголы, связанные с движением ('выйти', 'слезать с лошади', 'войти', 'утонуть')
- Низкая переходность, низкая предсказуемость:
 - Многие глаголы эмоций и чувств ('сердиться', 'радоваться', 'раздражаться', 'удивляться', 'доверять', 'влюбиться')

Предсказуемость моделей: ВЫВОДЫ

- Есть некоторые группы значений, которые обычно кластеризуются в рамках валентностных классов:
 - симметричные предикаты,
 - глаголы, связанные с обладанием,
 - и т.д...
- Никакой класс не соизмерим по степени межъязыковой регулярности с классом семантически высоко переходных глаголов.

Сравнивая глаголы: выводы

- Идея постулировать новую семантическую роль каждый раз, когда обнаруживается язык, который проводит соответствующее противопоставление в грамматике, привела бы самое это понятие к обесмысливанию (ролей стало бы бесконечно много).
- Валентностные свойства глаголов с более конкретной семантикой кросс-лингвистически более предсказуемы, чем валентностные свойства глаголов с более абстрактной семантикой.

Сравнивая глаголы: выводы

- Некоторые общепризнанные семантические роли (е.g. глагол с Экспериенцером и Стимулом) очень плохо предсказывают валентностные свойства глаголов.
- Полученные данные скорее не подтверждают идею об относительной атомарности семантических ролей, т.е. идею о том, что глагольные значения «строятся» из комбинаций ролей..
- Скорее, холистические глагольные значения оказываются первичными по отношению к семантическим ролям.

Сравнивая языки

Подробнее см. [Say 2014]

Доля (не)переходности

- В каждом языке сколько-то из 130 значений выражается переходными структурами, а сколько-то - непереходными
- Можно высчитать долю.

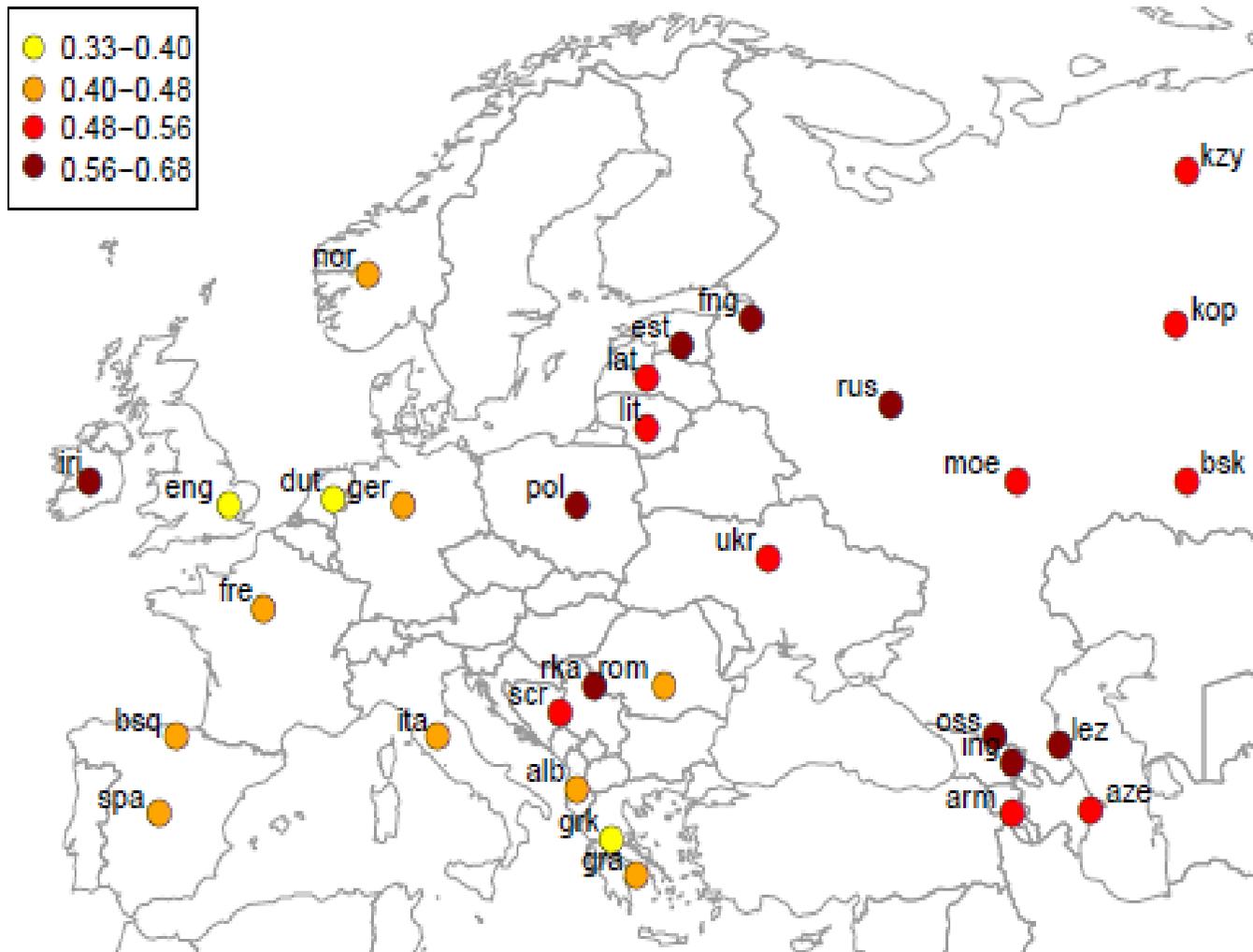
Доля непереходности

ингушский	67%	коми-пермяцкий	54%	немецкий	42%
эстонский	66%	эрзянский	53%	испанский	41%
лезгинский	66%	тувинский	53%	румынский	41%
ингерманландский	63%	багвалинский	52%	калмыцкий	41%
осетинский	61%	азербайджанский	52%	чукотский	41%
цыганский (кэлд.)	60%	латышский, сербский	50%	нидерландский	39%
русский	59%	армянский	50%	арабский	39%
польский	58%	албанский	48%	баскский	38%
гуро	58%	японский	46%	французский	37%
ирландский	57%	норвежский	46%	английский	37%
украинский	56%	цахурский	45%	гуарани	35%
коми-зырянский	55%	древнегреческий	44%	новогреческий	33%
башкирский	54%	лоома	44%	кхмерский	26%
литовский	54%	итальянский	42%	китайский	19%

Доля непереходности

- Большой разброс: от 0.19 до 0.67.
- Эти числа нельзя, конечно, воспринимать как абсолютное свойство языка: они зависят от (произвольного) выбора глаголов
- Однако сравнение языков может обнаруживать неслучайные различия и закономерности

Доля непереходности*



*Все карты выполнены в R Машей Овсянниковой

Доля непереходности

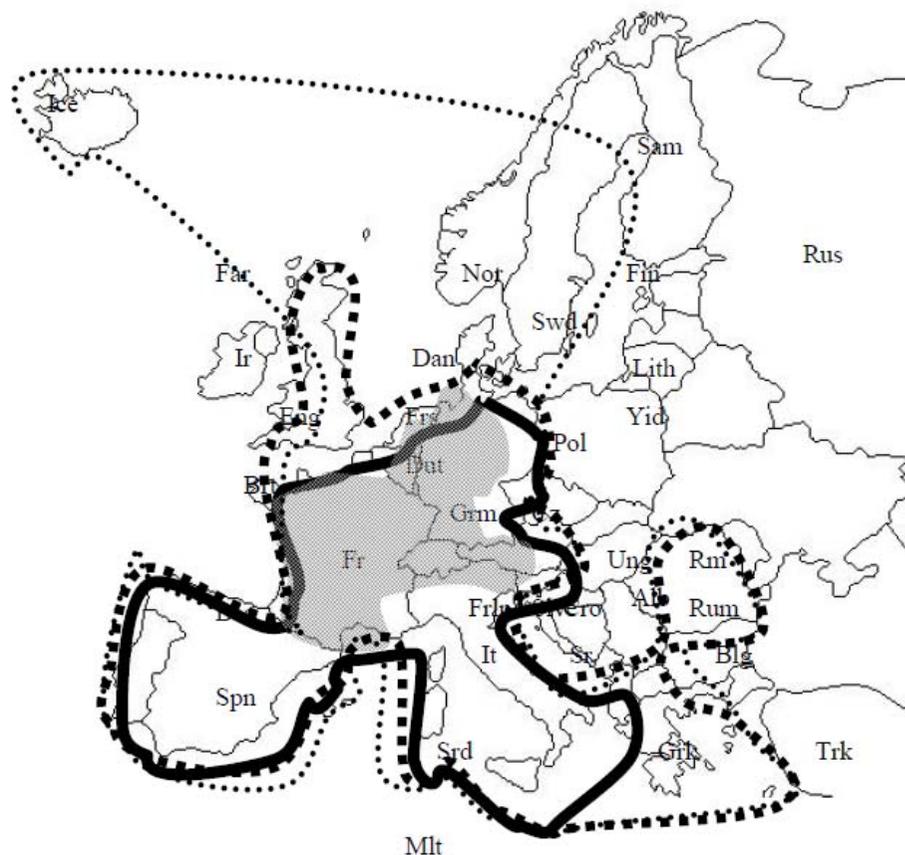
- Среди рассмотренных европейских языков самая высокая доля переходных в романских, германских, балканских и баскском.
- Это напоминает зону «Standard Average European» (см. Уорф, Хаспельмат и др.).

Standard Average European

[Haspelmath 2001]: the twelve SAE features:

- definite and indefinite articles (e.g. English *the* vs. *a*);
- postnominal relative clauses with inflected relative pronouns;
- 'have' + passive participle perfects;
- a preponderance of generalizing predicates to encode experiencers (\approx экспериенцеры-подлежащие);
- a passive construction formed with a passive participle plus an intransitive copula-like verb;
- a prominence of anticausatives in inchoative-causative pairs;
- dative external possessors;
- verbal negation with a negative indefinite;
- particle comparatives in comparisons of inequality;
- equative constructions based on adverbial-relative clause structures;
- subject person affixes as strict agreement markers;
- differentiation between intensifiers and reflexive pronouns (e.g. German intensifier *selbst* vs. reflexive *sich*).

Standard Average European



Данные из [Haspelmath 2001: 1505]; карта из [Van der Auwera 2011].

- 9 features: shading,
- 8 features **—————**
- 7 features **.....**
- 6 features **.....**

Standard Average European

[Haspelmath 2001]: the twelve SAE features:

- definite and indefinite articles (e.g. English the vs. a);
- postnominal relative clauses with inflected relative pronouns;
- 'have' + passive participle perfects;
- **a preponderance of generalizing predicates to encode experiencers (≈ экспериенцеры-подлежащие);**
- a passive construction formed with a passive participle plus an intransitive copula-like verb;
- a prominence of anticausatives in inchoative-causative pairs;
- dative external possessors;
- verbal negation with a negative indefinite;
- particle comparatives in comparisons of inequality;
- equative constructions based on adverbial-relative clause structures;
- subject person affixes as strict agreement markers;
- differentiation between intensifiers and reflexive pronouns (e.g. German intensifier selbst vs. reflexive sich).

Доля переходности и SAE

- Получается, что наши данные еще раз подтверждают уже известное?
- НЕ совсем: оппозиция между “обобщающими” и “инвертированными” экспериенциальными предикатами связана с тем, кодируется ли экспериенцер как подлежащее, но не с переходностью (cf. Bossong 1998; Haspelmath 2001: 1495ff.).

Доля переходности и SAE

- Присутствие неканонических A-аргументов (включая неподлежащие экспериенцеры) в наших данных не настолько велико, чтобы объяснить высокой уровень непереходности периферийных языков Европы
- Если рассматривать географическое распределение только для неканонических O, то получится примерно такая же карта.

Профили языков по (не)переходности

- Можно сравнивать пары языков по наборам тех предикатов, которые в них оказываются (не)переходными
- Похожие доли переходных предикатов логически не гарантируют низкое расстояние Хэмминга:

		ЯПОНСКИЙ	
		vt	vi
ЛАТЫШСКИЙ	vt	36	11
	vi	15	36

Профили языков по (не)переходности

- Слова расстояния Хэмминга (теперь между языками).
- NeighborNet (та же программа).

Профили языков по (не)переходности

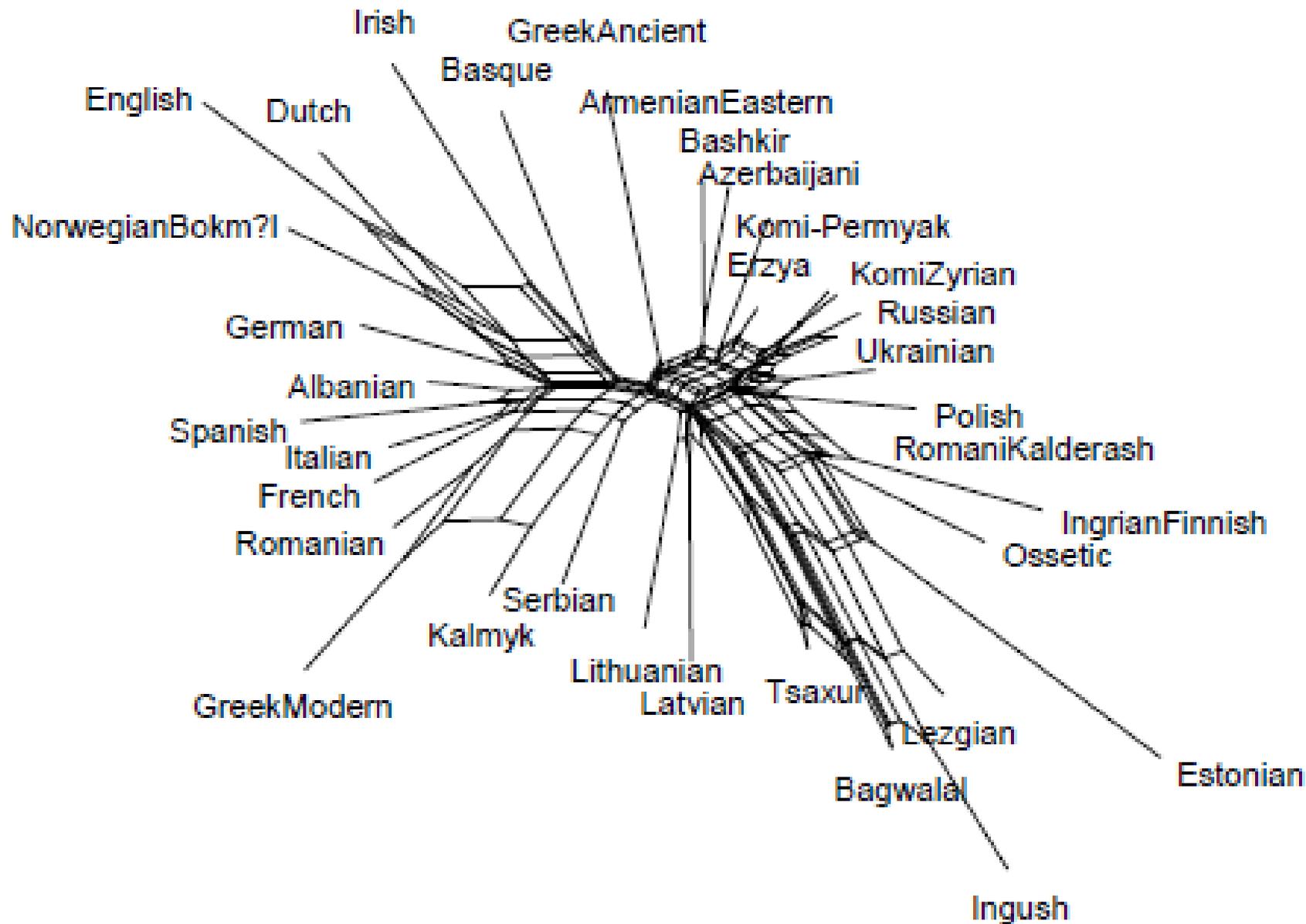
- Картина снова получилась почти одномерной: от “более переходных” к “менее переходным” языкам.
- Это означает, что степень переходные классы в менее переходных языках хорошо вкладываются в аналогичные классы в более переходных языках.
- Еще раз подтверждает существование **иерархии** вероятности переходности у глагольных значений.

Профили языков по (не)переходности: Европа

- Как структурируются языки Европы по их «профилям» (не)переходности?
- Не будем учитывать глаголы, для которых много пробелов (20% и больше).
- Заметны ареальные схождения! См., например, положение ирландского (ср. выше о доле переходности в языках SAE и за пределами этого языкового союза).

Профили (не)переходности: Европа

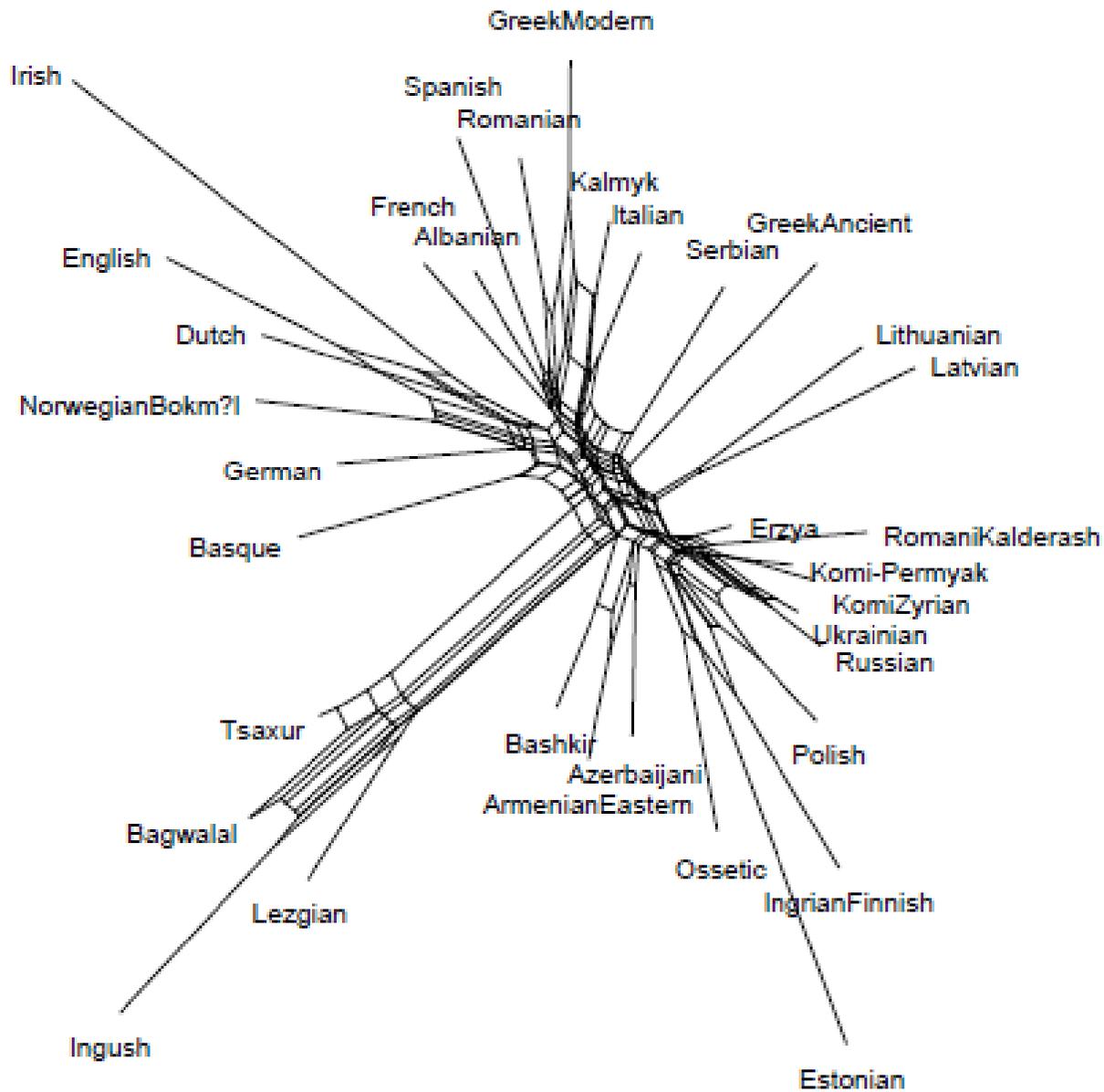
0.01



Профили языков по «локусу»: Европа

- Та же техника, что до этого, но в качестве основы для расстояния используется относительная доля глаголов с несовпадающим признаком локуса в парах языков

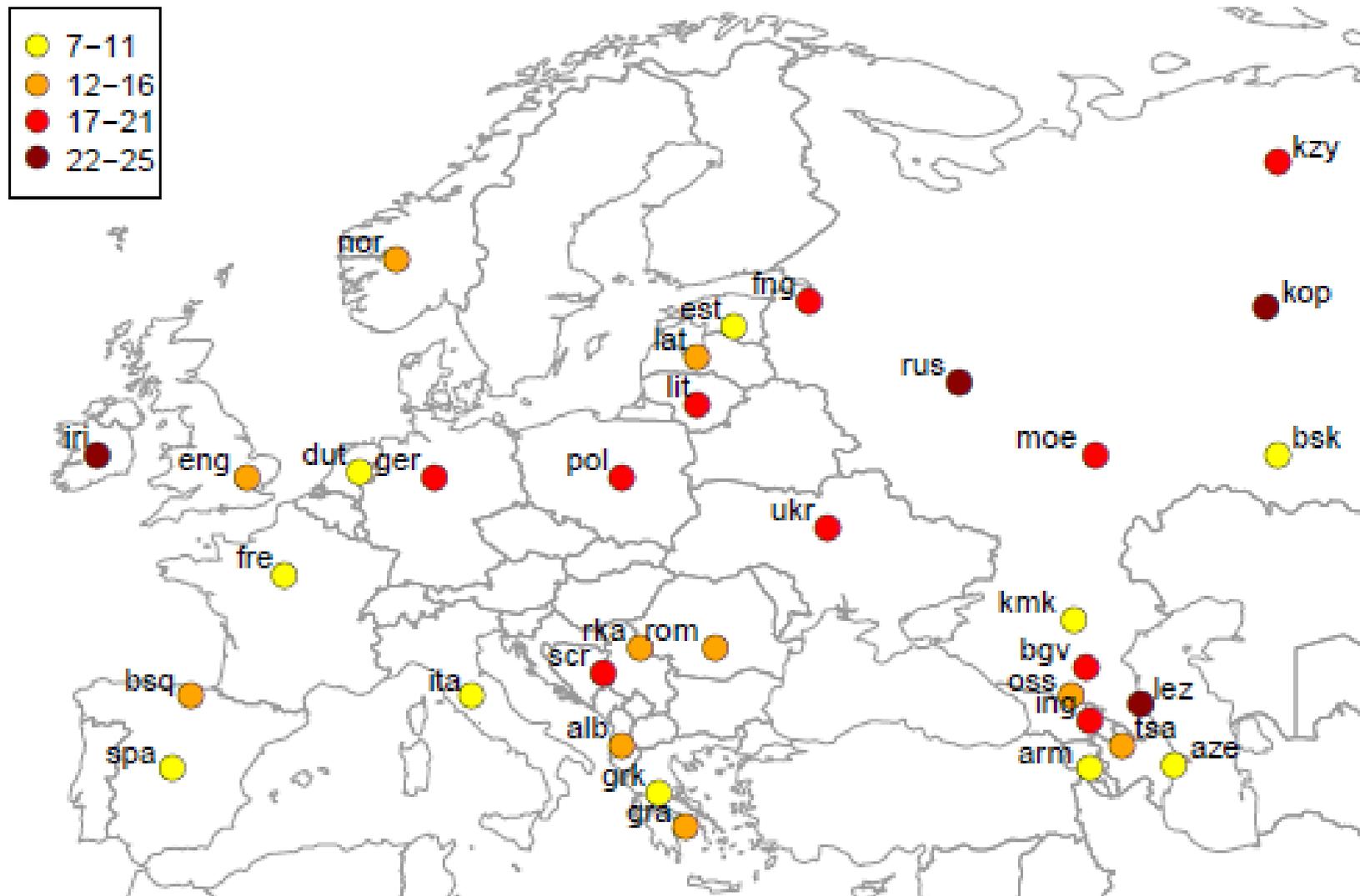
Профили языков по локусу: Европа



Количество валентностных классов

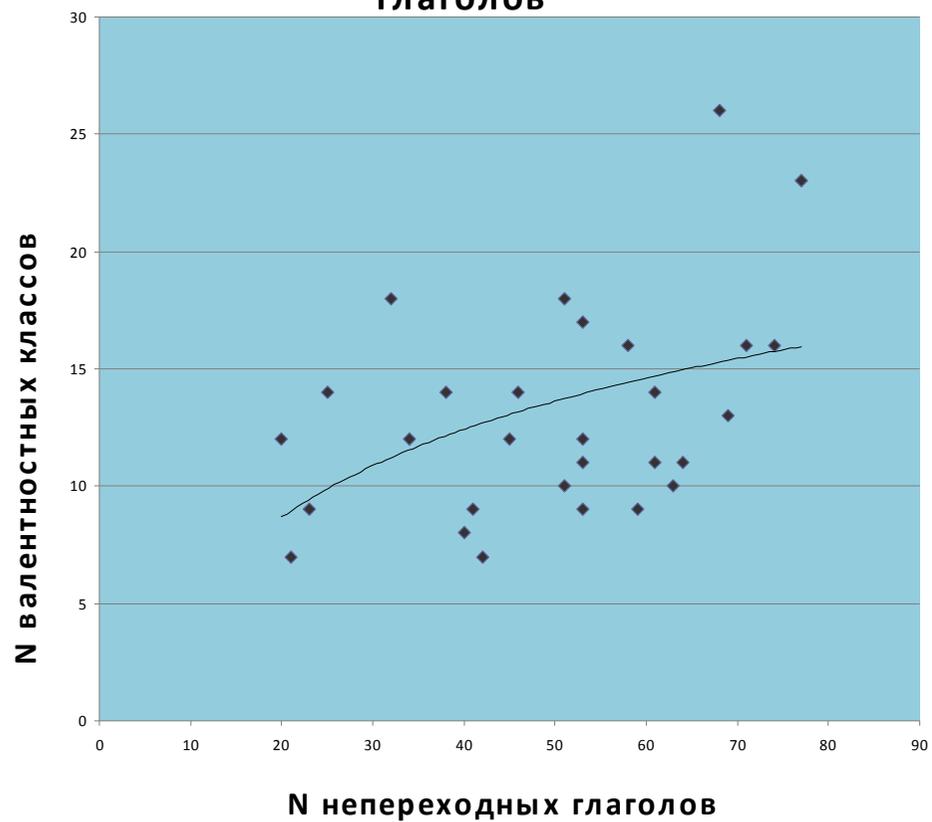
- Все переходные глаголы похожи друга на друга, каждый непереходный глагол непереходен по-своему.
- Как тогда типологически изучать непереходные классы?
- **Количество** глагольных классов?

Количество валентностных классов



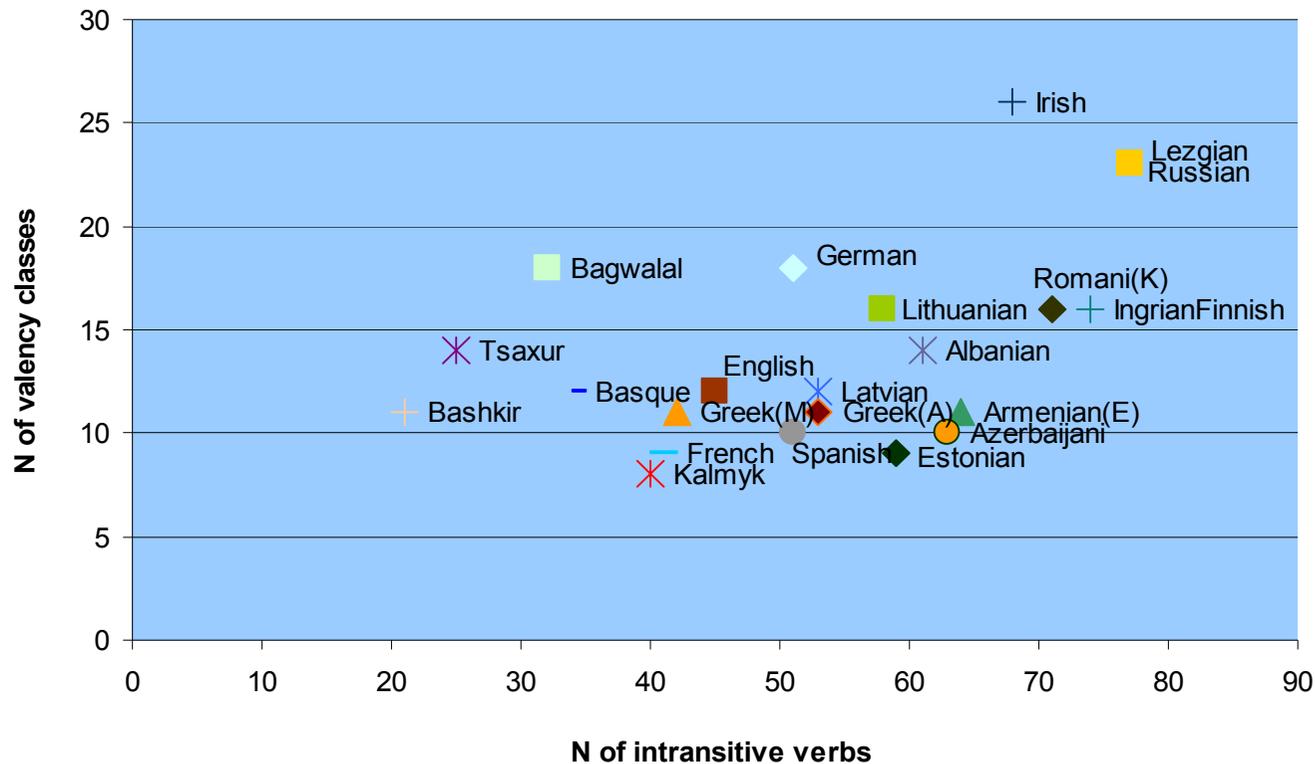
Количество валентностных классов

... и количество непереходных глаголов



Количество валентных классов

Количество непереходных глаголов и количество глагольных классов: языки Европы



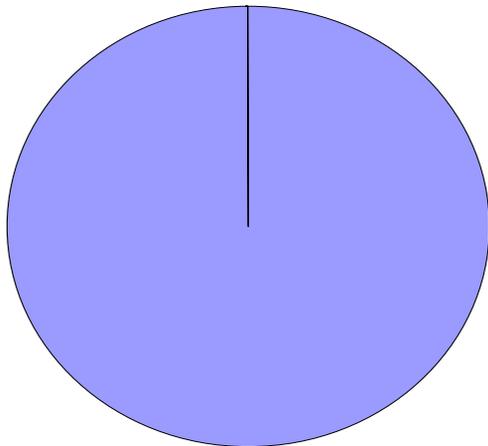
Количество валентностных классов

Промежуточные выводы:

- нет четкого географического распределения,
- очень сильно зависит от количества зафиксированных непереходных глаголов.
- Не самая лучшая мера при избранной методике. Что вместо этого?
- **Энтропия:** «количество беспорядка».

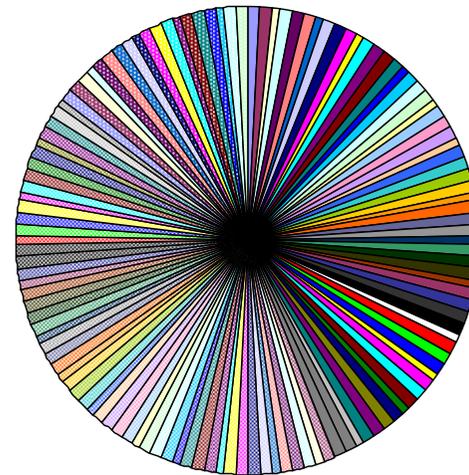
Сравнивая языки: энтропия

$$H(x) = - \sum_{i=1}^k p(x_i) \cdot \log(p(x_i))$$



Гипотетический язык №1:
Все глаголы попали в один класс

$$H = 0$$

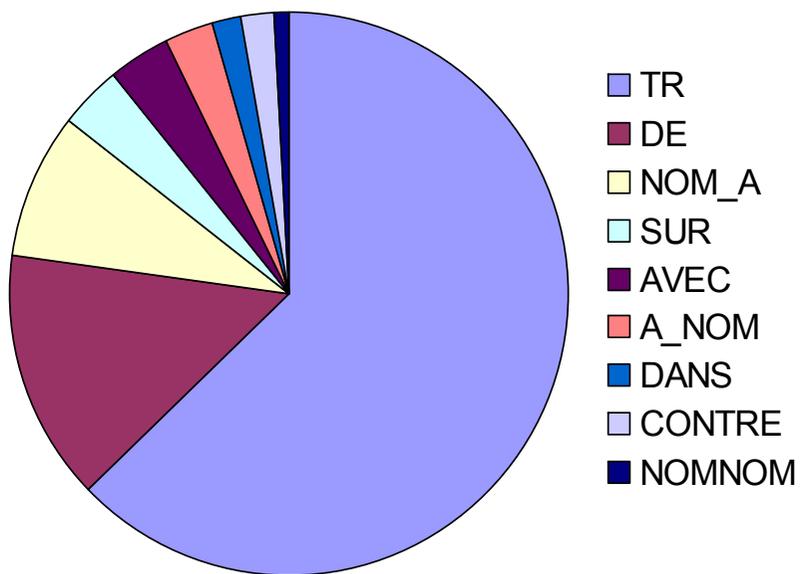


Гипотетический язык № 2:
130 глагольных классов

$$H = -\log\left(\frac{1}{130}\right) \approx 4,87$$

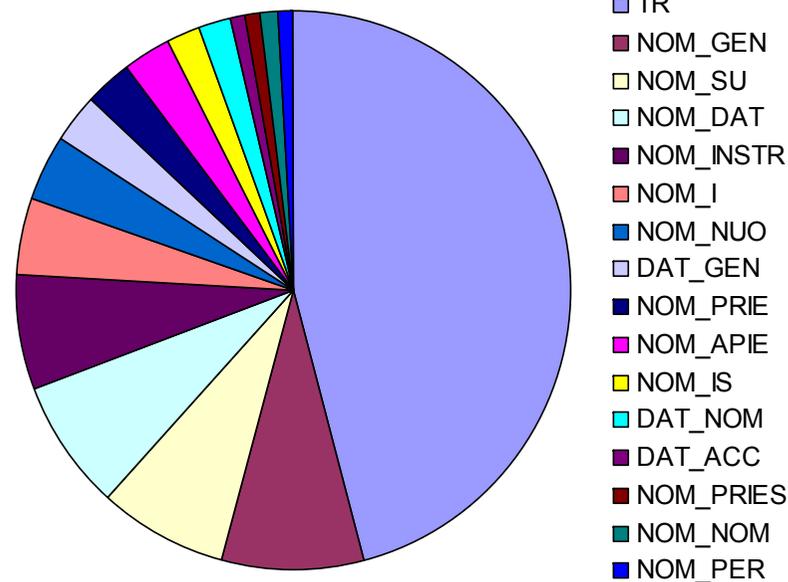
Энтропия

французский



$H \approx 1.31$

ЛИТОВСКИЙ



$H \approx 2.02$

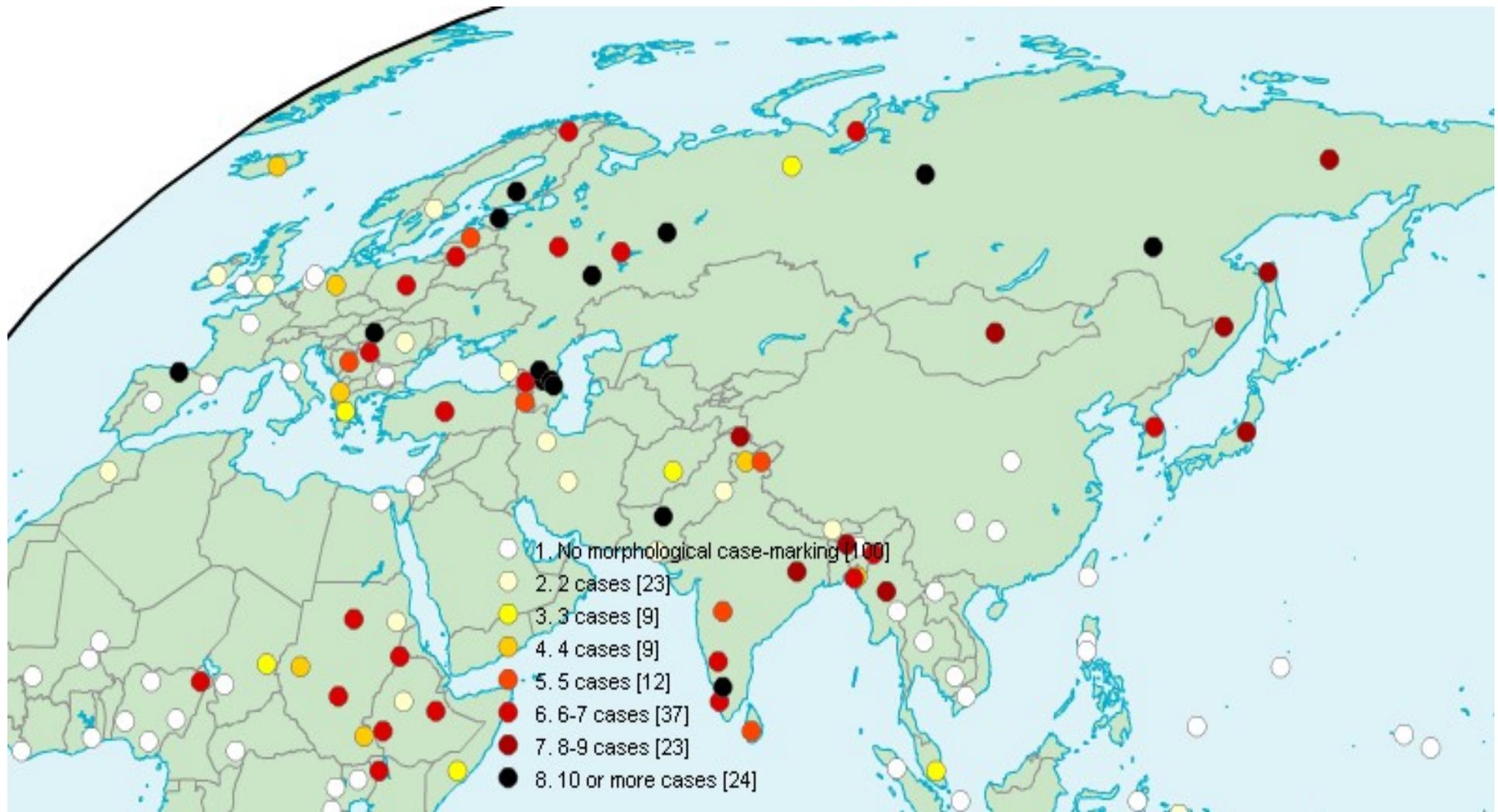
Энтропия

- На небольших выборках подсчет энтропии на основе наблюдаемых частот занижает реальную энтропию в популяции.
- Поэтому языки с меньшим количеством собранных глаголов демонстрируют меньшую энтропию.
- Относительная энтропия: $H/\ln(N_{\text{verbs}})$, где N_{verbs} – количество предикатов, для которых были собраны данные.
- По сути: наблюдаемая энтропия сравнивается с теоретическим максимумом.

Энтропия

- Энтропия распределения по классам тоже коррелирует с количеством непереходных глаголов.
- Однако эта мера имеет тот плюс, что она учитывает структуру непереходных классов.
- Снова ареальная тенденция: центральные языки зоны SAE имеют меньшую энтропию, чем языки европейской периферии.
- Видимо, коррелирует с количеством падежей.
- (И, возможно, с грамматикализованностью предлогов / послелогов? (Пока нет ясности))

Количество падежей (WALS)



Энтропия

- Примерно то же распределение получается, если учитывать только непереходные классы. Но этих данных мало.

Как типологически изучать малые валентностные классы?

- Можно ли опираться на дескриптивные грамматические ярлыки в отдельных грамматиках («аблатив», «косвенное дополнение»)?
- НЕТ! Одни и те же ярлыки могут обозначать совсем разные вещи (и эта проблема неразрешима).
- Например, в балто-финских грамматиках «генитивом» называют ядерный падеж Р-участника в базовой переходной конструкции.

Как типологически изучать малые валентностные классы?

- Опирается на отдельные пары значений.
- Возможно. Например, можно построить частную типологию по такому признаку: попадают ли ‘зависеть’ и ‘отличаться’ в один класс.

Один и тот же класс	Разные классы
русский, эстонский, ингерманландский, французский, немецкий, литовский, латышский, лезгинский, новогреческий, албанский, испанский, армянский, азербайджанский... (+гуро)	английский, др.-греческий, башкирский, ирландский, цыгаснский (кэлд.)... +арабский, японский, кхмерский, китайский

Как типологически изучать малые валентностные классы?

- По сути это подход, известный в области грамматики как метод «семантической карты».
- Однако при помощи такого метода сложно увидеть целостную картину – пар лексических значений слишком много (больше чем на привычных семантических картах в грамматике).

Как типологически изучать малые валентностные классы?

- Альтернатива: строить матрицу соответствий глагольных классов в парах языков. Например, древнегреческий (по горизонтали) и новогреческий (по вертикали).

	TR	GEN	N_DAT	DAT_N	eis	uper	dia	pros	eks	apo	en
TR	58	11	8		3	1					
me	6		7				1	1			
apo		2	3						2	1	
se		1	3		1			1			1
GEN_N	2	1		2							
gia			1								

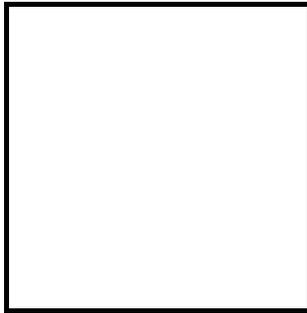
	TR	GEN	N_DAT	DAT_N	eis	uper	dia	pros	eks	apo	en
TR	58	11	8		3	1					
me	6		7				1	1			
apo		2	3						2	1	
se		1	3		1			1			1
GEN_N	2	1		2							
gia			1								

	TR	GEN	N_DAT	DAT_N	eis	uper	dia	pros	eks	apo	en
TR	58	11	8		3	1					
me	6		7				1	1			
apo		2	3						2	1	
se		1	3		1			1			1
GEN_N	2	1		2							
gia			1								

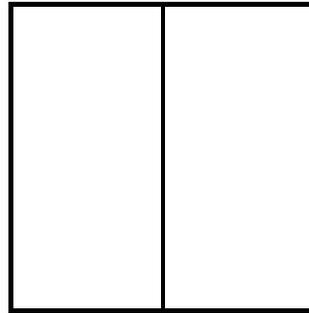
Как типологически изучать малые валентностные классы?

- Как измерить степень похожести систем?

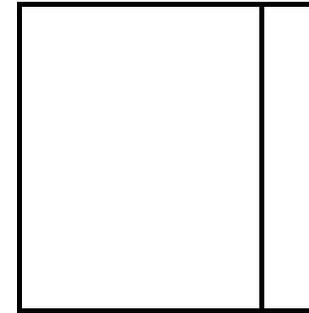
Entropy



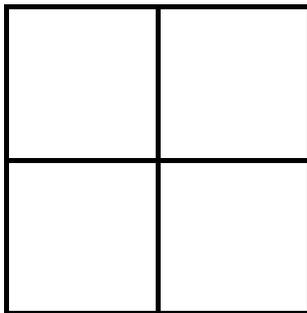
$$H = 0$$



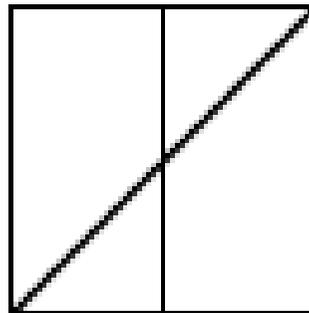
$$H \approx 0,69$$



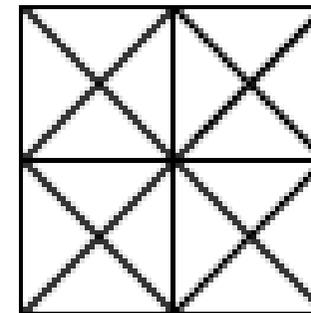
$$H \approx 0,50$$



$$H \approx 1,39$$

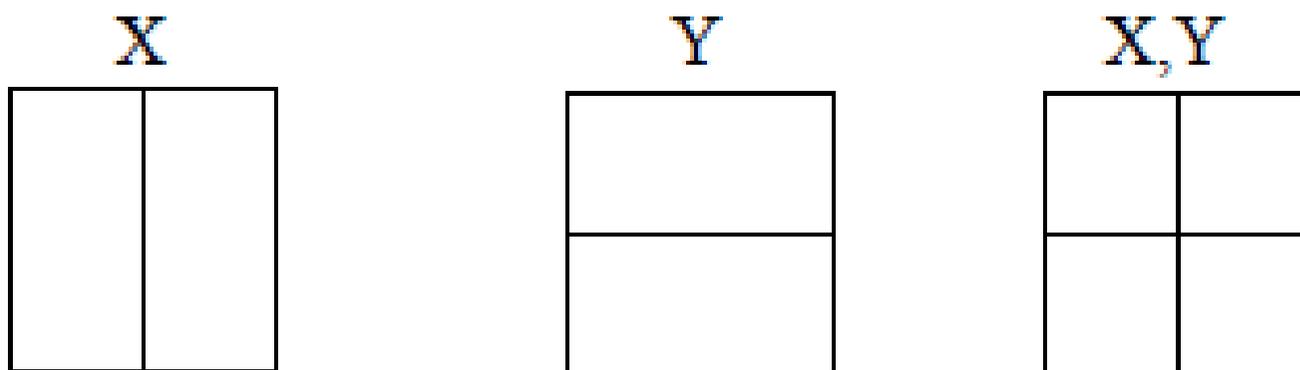


$$H \approx 1,26$$



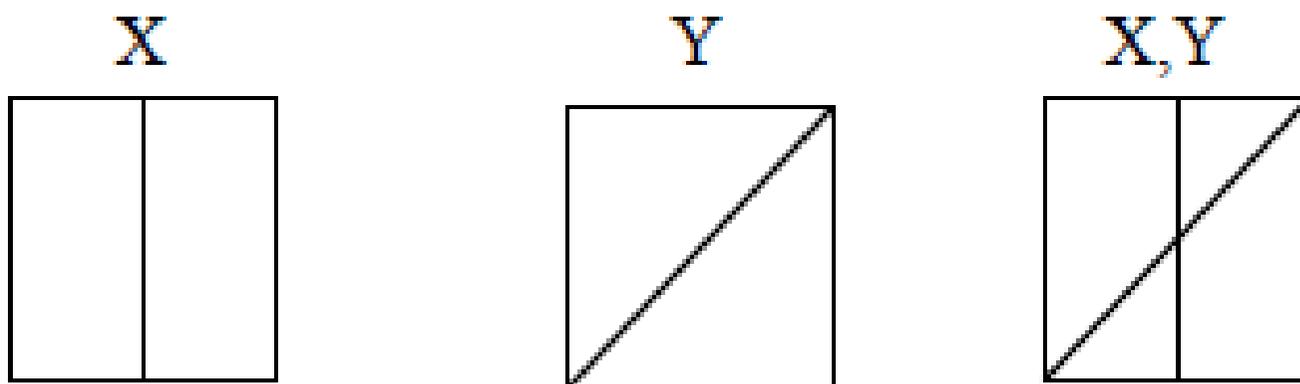
$$H \approx 2,77$$

Mutual information



$$I(X; Y) = H(X) + H(Y) - H(X, Y) \approx 0,69 + 0,69 - 1,39 = 0$$

* * *



$$I(X; Y) = H(X) + H(Y) - H(X, Y) \approx 0,69 + 0,69 - 1,26 \approx 0,13$$

Predictability

X

Y

X,Y

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \approx 0,69 + 1,39 - 1,39 \approx 0,69$$

Predictability of X given Y :

$$\pi(X|Y) = I(X;Y) / H(X) = 0,69 / 0,69 = 1$$

Predictability of Y given X :

$$\pi(Y|X) = I(X;Y) / H(Y) = 0,69 / 1,39 = 0,5$$

Как типологически изучать малые валентностные классы?

- Таким образом, у нас появляется возможность измерить похожесть (взаимную информацию, MI) систем глагольных классов в любой паре языков.
- Рассчитаем расстояния в парах языков:

$$D(X, Y) = 1 - \frac{\pi(X|Y) + \pi(Y|X)}{2}$$

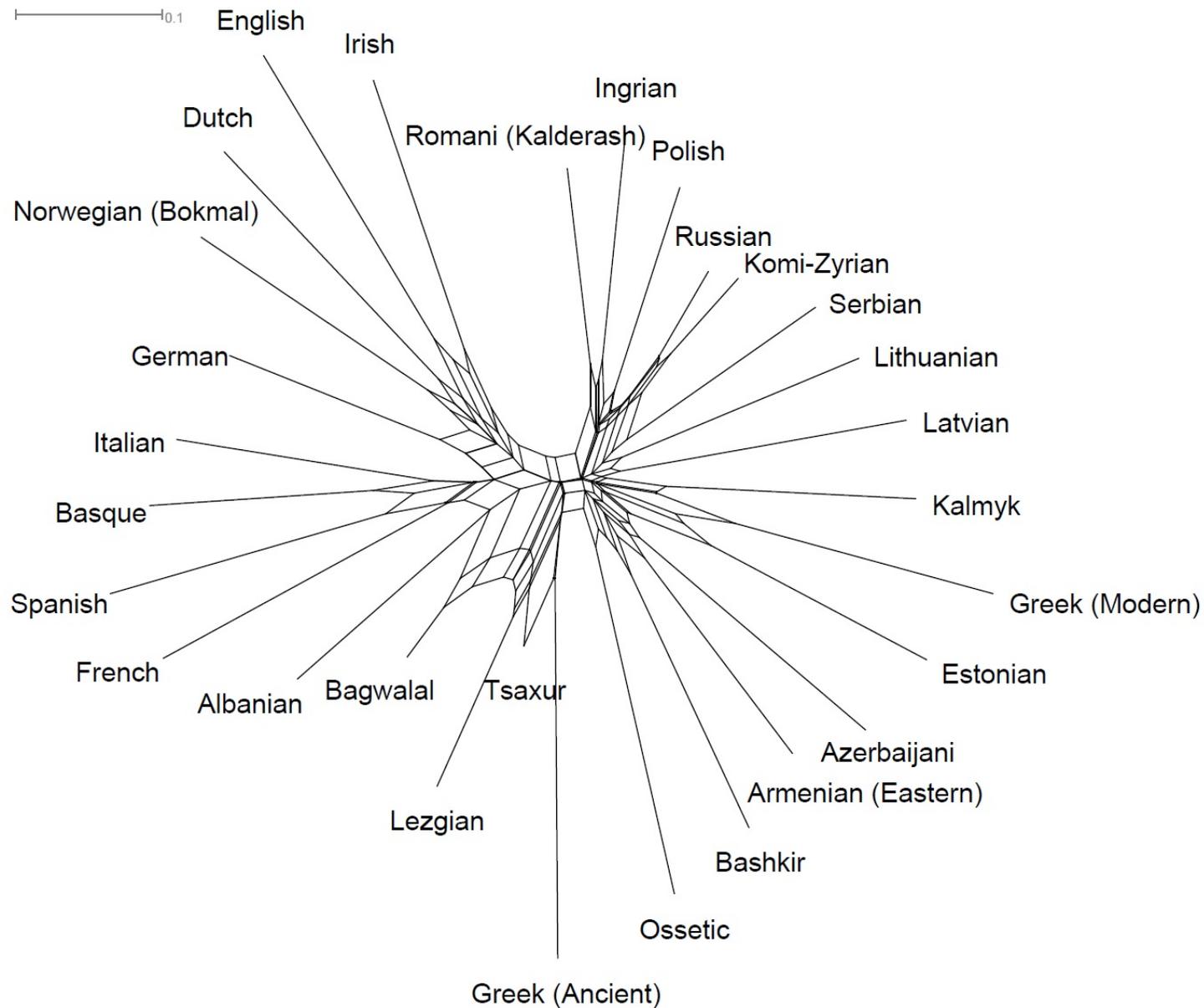
- Мы можем построить матрицу для всего множества изученных языков.
- И использовать все тот же SplitsTree для визуализации: построения NeighborNet.

NeighborNet для множества языков на основе предсказуемости классов

- удаляем неевропейские языки,
- удаляем языки, для которых мало данных (цахурский и багвалинский),
- удаляем языки, для которых данные были собраны по опубликованным источникам,
- удаляем языки, которые я не успел учесть :)
- остается 25 языков

NeighborNet, основанный на взаимной предсказуемости классов в парах языков

25 языков Европы



Выводы

- Общая склонность к (не)переходности и общая сложность системы валентностных классов, видимо, демонстрируют ареальные закономерности большого масштаба (например, SAE vs. языки европейской периферии)

Выводы

- Сходства в структуре индивидуальных классов, вероятно, больше отражает конкретные контакты и неглубокие генетические связи
- Если это правда, то это косвенное свидетельство в пользу того, что системы валентностных классов могут меняться сравнительно быстро
- Видимо, быстрее, чем индивидуальные морфосинтаксические средства, используемые для кодирования (ср. данные для древнегреческого и новогреческого).

Спасибо!



Избранная библиография

- Апресян, Ю.Д. 1967. Экспериментальное исследование семантики русского глагола. М.: Наука.
- Aikhenvald A.Y., Dixon R.M.W., Onishi M. (eds). 2001. Non-canonical marking of subjects and objects. Amsterdam/Philadelphia: John Benjamins.
- Bhaskararao P., Subbarao K.V. (eds). 2004. Non-nominative Subjects. 2 vols. Amsterdam/Philadelphia: John Benjamins.
- Bickel, B. 2010. Quantitative Methods in Typology. Course taught at DGfS-CNRS Summer School on Linguistic Typology, Leipzig, August 15 - 27 [<http://www.uni-leipzig.de/~bickel/lehre/Leipzig2010/>]
- Bickel, Balthasar, Alena Witzlack-Makarevich, Taras Zakharko & Lennart Bierkandt. 2011. Semantic role clustering: an empirical assessments of semantic role types. A talk given in Zurich (??).
- Bossong, Georg. 1998. Le marquage de l'expérience dans les langues d'Europe'. In: Feuillet (ed.). Actance et valence dans les langues de l'Europe. Berlin: Mouton de Gruyter. 259–94.
- Dixon, R.M.W. 1994. Ergativity. Cambridge: CUP.
- Dixon, R.M.W., Aikhenvald, A.Y. (eds). 2000: Changing valency: case studies in transitivity. Cambridge.
- Dowty D. 1991. Thematic proto-roles and argument selection. Language, 67. 547-619.

Избранная библиография

- Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages // Aikhenvald, Alexandra Y., Robert M. W. Dixon, and Masayuki Onishi. (eds.). 2001. Non-canonical marking of Subjects and Objects. [=Typological studies in language, 46]. Amsterdam: John Benjamins. 53-83.
- Hopper, P.J., Thompson, S.A. 1980. Transitivity in grammar and discourse. *Language*. 1980, 56. (2). P. 251–299.
- Kittilä, S. 2002. Transitivity: towards a comprehensive typology. Turku, 2002.
- Levin, Beth. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- Malchukov, A. 2006. Transitivity parameters and transitivity alternations: constraining co-variation. In: *Case, valency and transitivity*, ed. by L. Kulikov, A. Malchukov, P. de Swart. Amsterdam, Philadelphia. P. 175–190.
- Næss, Å. 2007. *Prototypical Transitivity*. Amsterdam, Philadelphia.
- Say, S. 2014. Bivalent Verb Classes in the Languages of Europe: A Quantitative Typological Study. *Language dynamics and change*, 4 (2014), 116–166.
- Tsunoda, T. 1981. Split case-marking patterns in verb-types and tense / aspect / mood // *Linguistics*. Vol. 19. P. 389–438.
- Van Belle, William & Willy van Langendonck (eds.). 1996. *The dative*. [= Case and grammatical relations across languages, Vol. 2]. Amsterdam / Philadelphia: John Benjamins.
- Yip, M., Maling, J. & Jackendoff, R. 1987. Case in tiers. *Language* 63 (2). 217–250.