

Данные в типологии

Сергей Сай
serjozhka@yahoo.com

**= Типы данных в типологии,
их (квантитативная) обработка и интерпретация**

1. Что такое типология?

- «Типология лингвистическая — сравнительное изучение структурных и функциональных свойств языков независимо от характера генетических отношений между ними <...>
- Типологическое исследование может иметь различные, но взаимосвязанные цели:
 - констатацию структурных сходств и различий между языками (*инвентаризационная* типология);
 - интерпретацию систем языков в плане совместимости — несовместимости структурных характеристик и предпочтительных типов структурной сообразности как систем в целом, так и отдельных уровней языка (*импликационная* типология);
 - классификацию языков по определённым типам и классам (*таксономическая* типология), которая обычно считается основной и конечной целью типологии».

Что такое типология?

- Примерно в таком виде типология
 - представлена в обязательных курсах на кафедре общего языкознания СПбГУ;
 - отражена в доступных учебниках, входящих в программы ряда теоретических курсов [Плунгян 2000 , Тестелец 2001, Comrie 1981].

Плунгян, В.А. 2000. Общая морфология. Введение в проблематику. М., Эдиториал УРСС.

Тестелец, Я. Г. 2001. Введение в общий синтаксис. М.: РГГУ. 2001.

Comrie, B. 1981. Language universals and linguistic typology. Chicago: University of Chicago Press.

Немного истории

Основные этапы развития типологии (ср. [Song 2011], но с модификациями):

- 3) 1980-1990-ые годы
 - поиск универсальных предпочтений (ранее известны как «фреквенталии»);
 - расцвет функционально-типологического направления: когнитивные объяснения;
 - стремление «очистить» данные от вневлингвистического шума (эффектов миграций, контактов и других исторических событий)
- **4) Современный этап: «популяционное» мышление (см. ниже)**

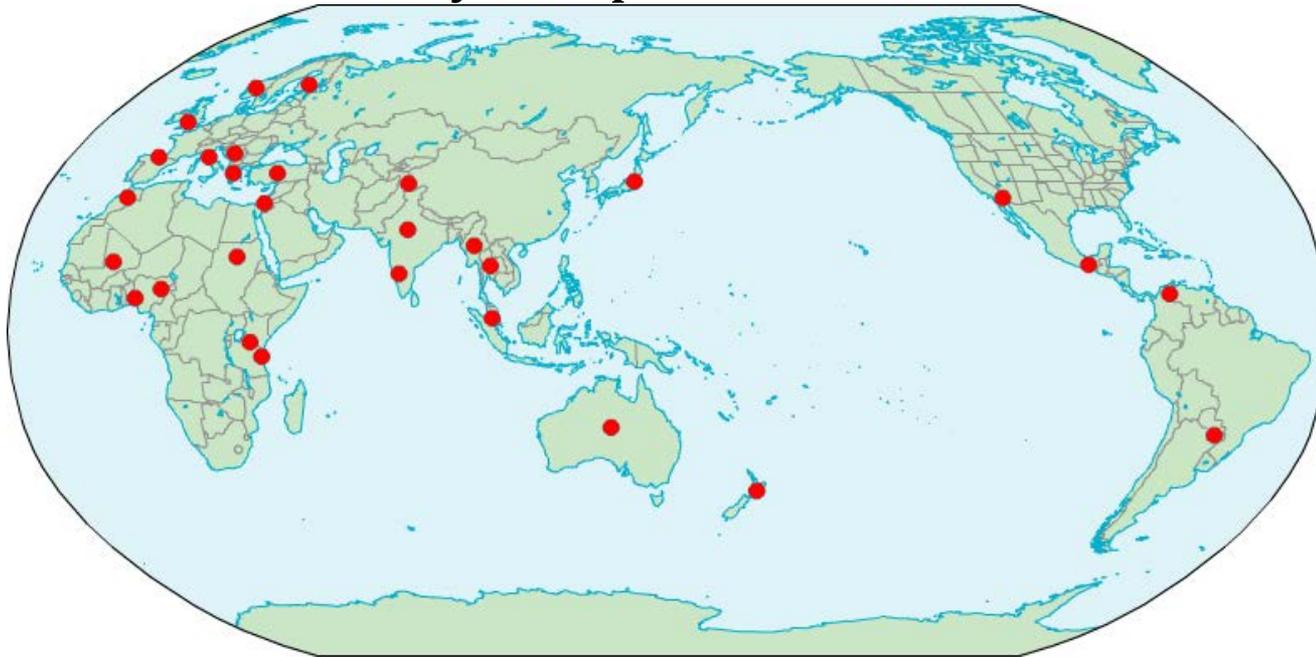
Современный этап: рост эмпирической базы

- Быстрый рост массивов данных:
 - вширь (количество рассматриваемых языков)
 - вглубь (сложность анализируемых структур)
- Основная проблема типологии 20 века:
 - найти достаточное количество надежных данных
- Основная проблема типологии 21 века:
 - обработать и проинтерпретировать то огромное количество данных, которые нам доступны (и успеть задокументировать незадокументированное)

Современный этап: рост эмпирической базы

Ср. выборку 30 языков, использованную в [Greenberg 1963]:

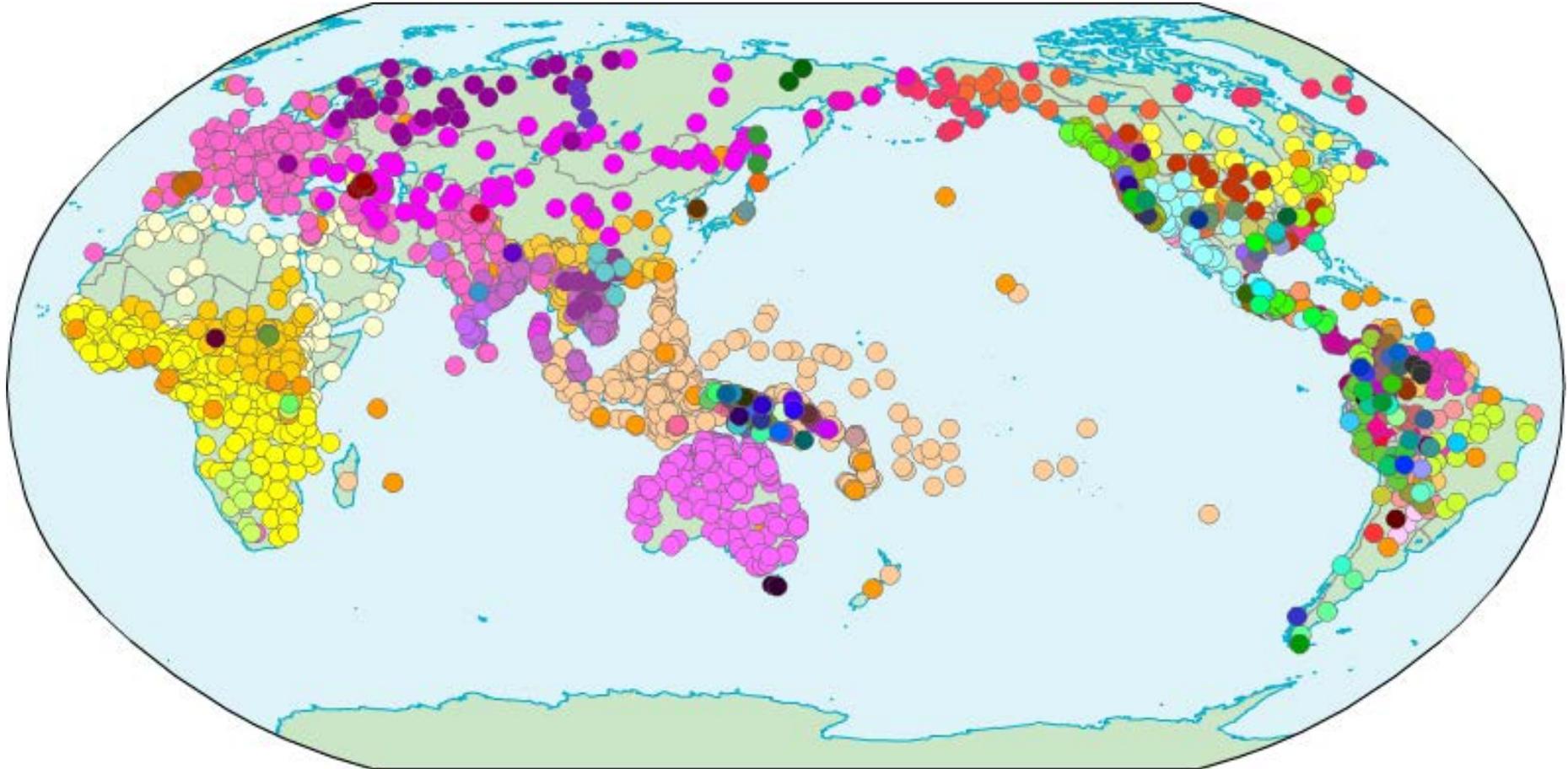
- До сих пор остается одним из основных источников гипотез о языковых универсалиях.



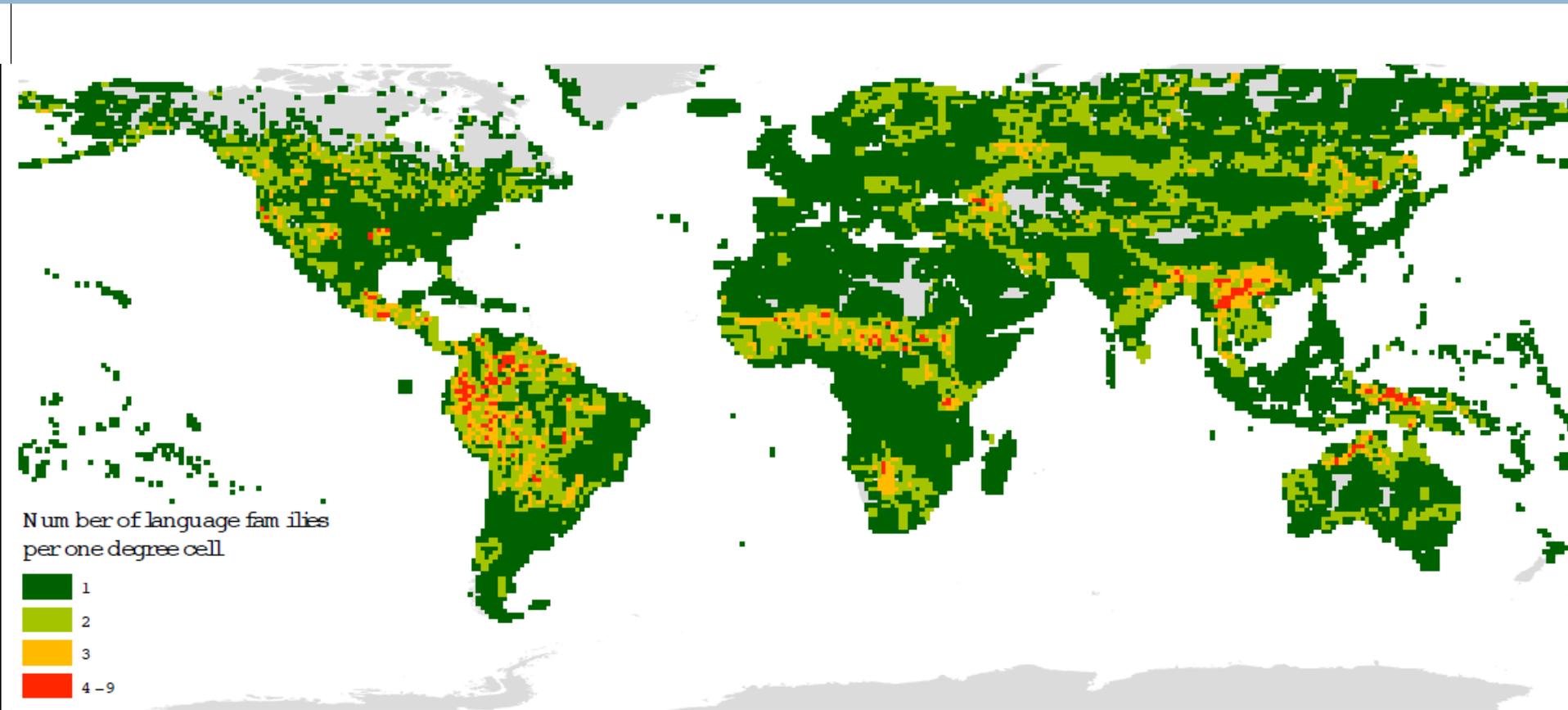
Greenberg, Joseph H. 1963. "Some universals of grammar with particular reference to the order of meaningful elements." In: Greenberg, Joseph H. (eds.) *Universals of grammar*, 73-113. Cambridge, Mass.: MIT Press.

Составление выборки

Языковые семьи (205) на карте мира: 2560 языков



Современный этап: рост эмпирической базы



Современный этап: рост эмпирической базы

- Не только увеличение количества, но и улучшение качества доступных грамматик
 - Взаимное влияние типологии и практики описания отдельных языков (“reference grammars”)
- Почти завершена работа по построению генетической классификации языков
 - Горизонт событий, доступных сравнительно-историческому методу, -- примерно 6000-8000 лет
- Достижения параллельных дисциплин в области ранней истории человечества / популяций:
 - генетика
 - антропология и этнография

Современный этап: методы и задачи

	20th century “Universalist Typology”	21st century “Distributional Typology”
nature	method	discipline
goal	formulating UG	explaining distributions
interest in	absolute implicational universals	distributional probabilities
part of	classical CogSci	anthropology, including psychology
core question	What’s a Possible/ Learnable Human Language?	What’s Where Why?

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/>]

Современный этап: методы и задачи

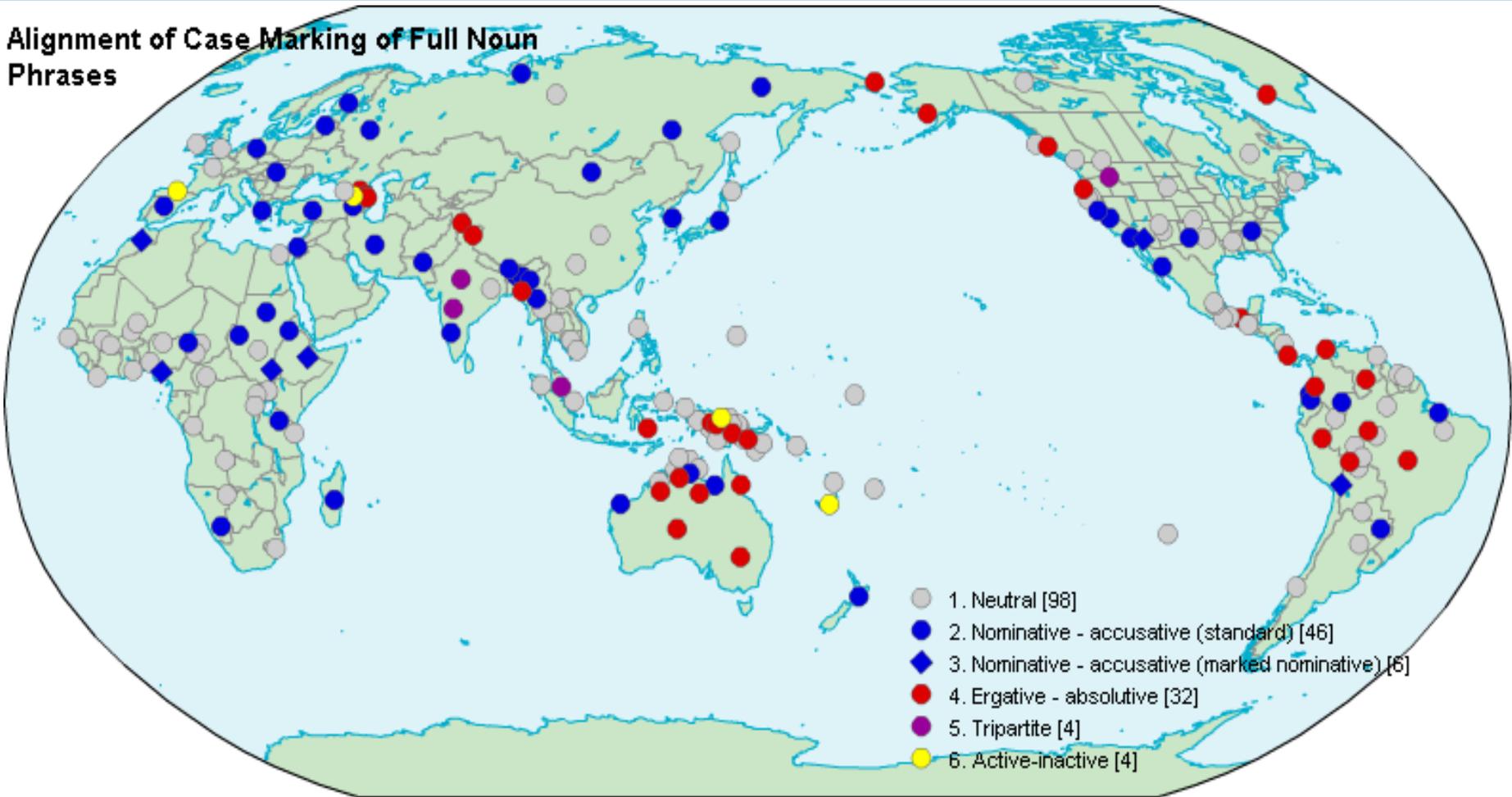
- Таким образом, типология встраивается в систему популяционных дисциплин, изучающих распределения:
 - генетика;
 - экология;
 - экономика;
 - антропология;
 - этнография (!);
 - фольклористика (!);
 - ...

План курса

1. Типологические параметры: какие типы данных используются в типологии?
2. Что типологические распределения говорят нам о языке?
3. Что типологические распределения говорят нам о языках?

Классические N-арные номинальные переменные

Alignment of Case Marking of Full Noun Phrases



«Этические» сети без фиксированного алфавита

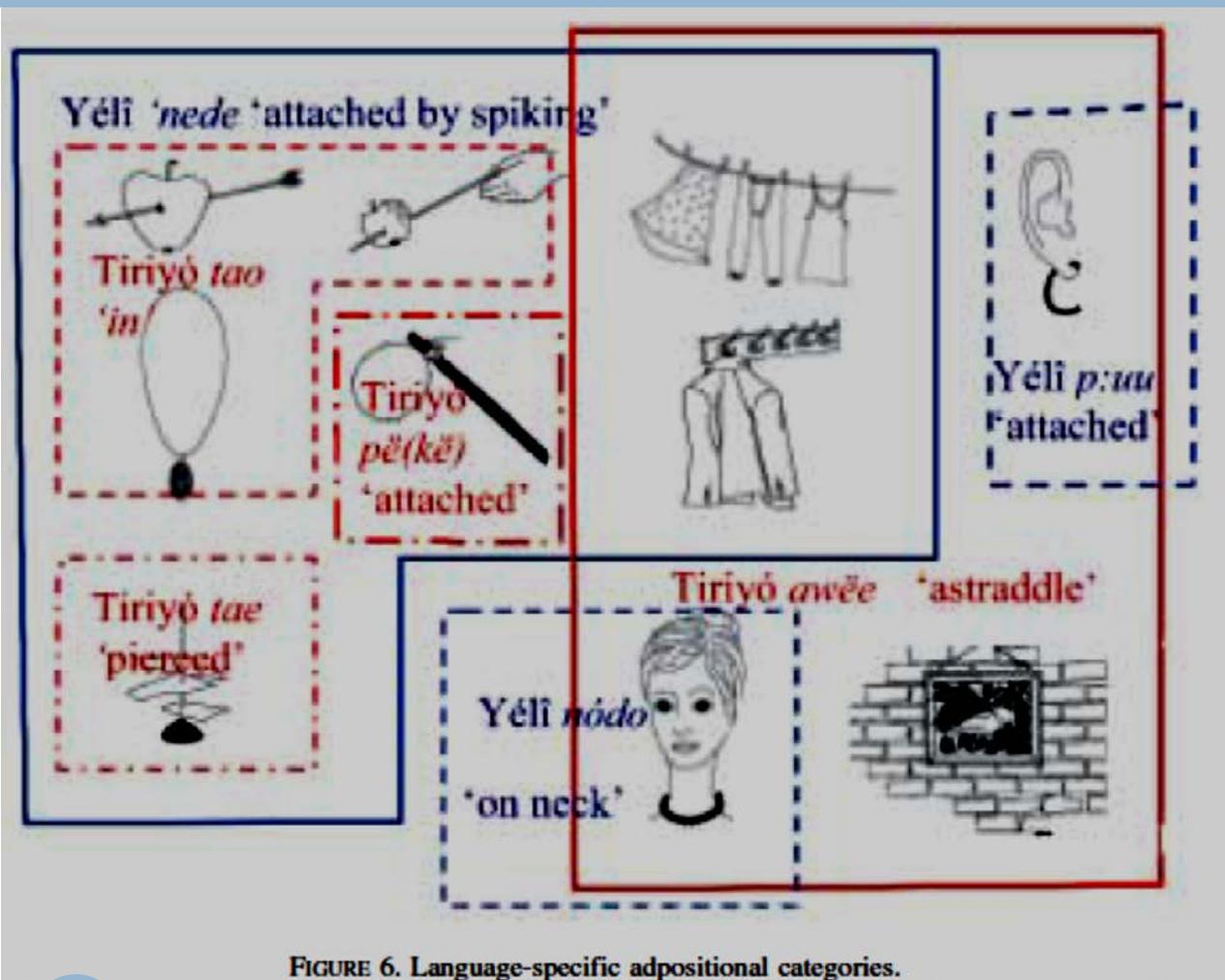


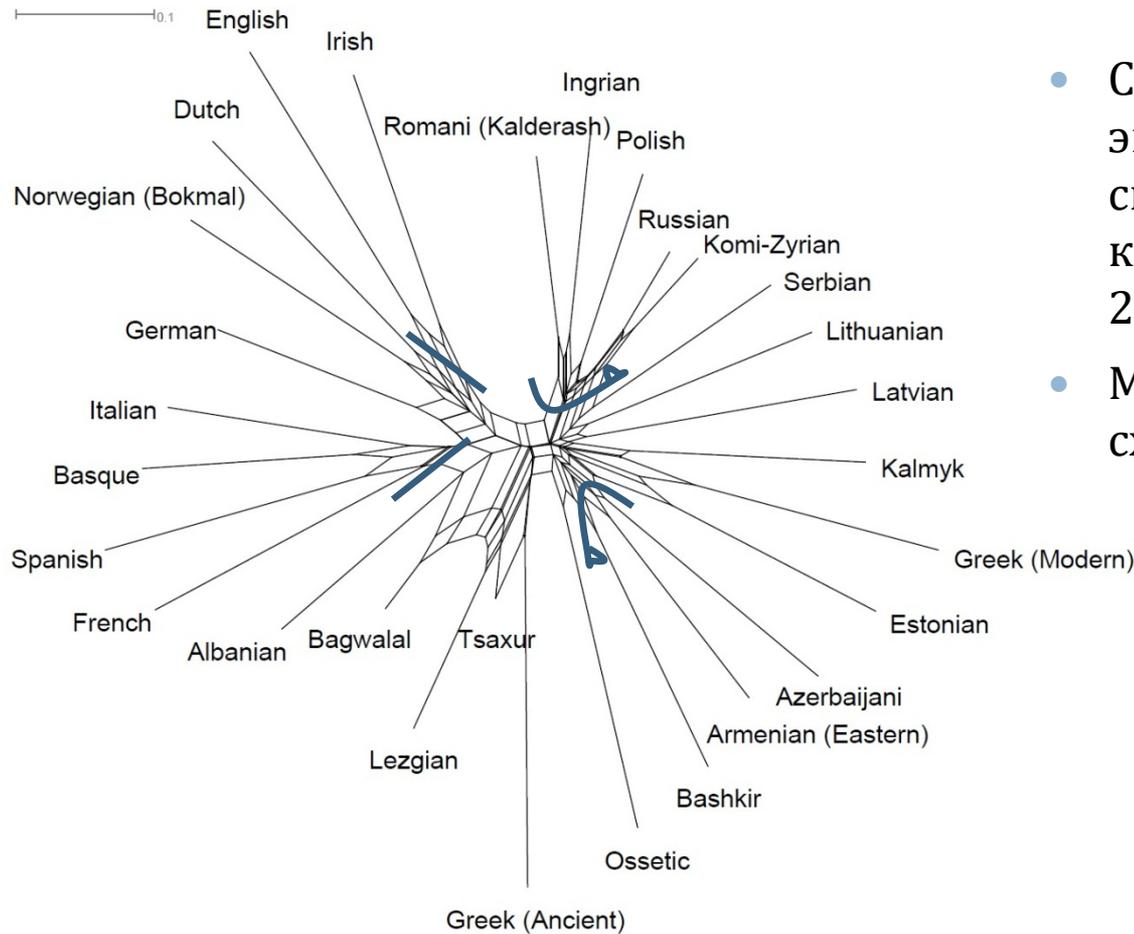
FIGURE 6. Language-specific adpositional categories.

Этическая сеть: 10 сценариев, представленных на рисунках

Эмические категории: предлоги/послелогои, используемые в языках тирийо (карибский, Бразилия) и йеле (йелидне, изолят, Папуа)

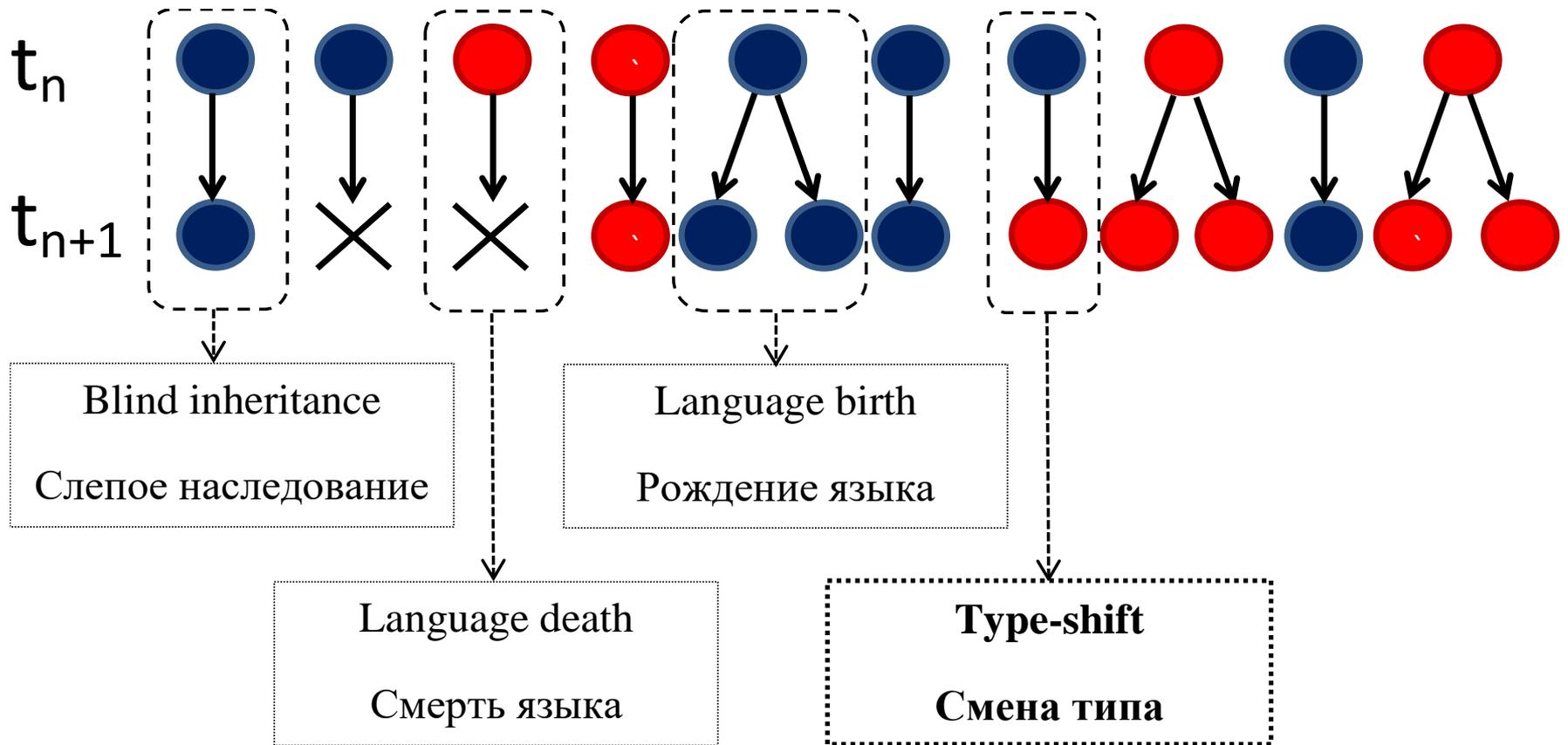
Levinson, Stephen, Sérgio Meira, The Language & Cognition Group. 2003. 'Natural Concepts' in the Spatial Topological Domain – Adpositional Meanings in Cross-linguistic Perspective: An Exercise in Semantic Typology. *Language* 79/3. P. 485-516.

Матрицы расстояний



- Сходства (измеренные через энтропию и MI) между системами валентностных классов в языках Европы [Say 2014]
- Множество микроареальных схождений!

Динамическая типология



Динамическая типология

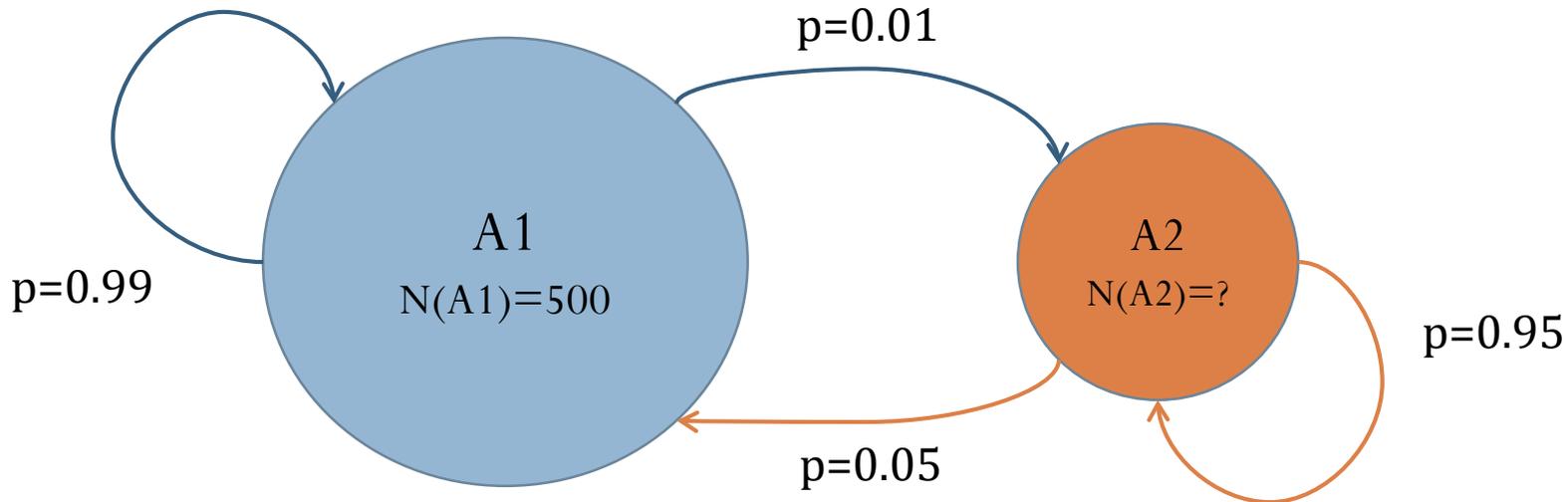
- Необходимые параметры модели:
 - $p(\text{Birth})$ – вероятность «расщепления» языка надвое с сохранением значения параметра
 - $p(\text{Death})$ – вероятность смерти языка
- Значения $p(B)$ и $p(D)$ определяют ожидаемое распределение размеров генетических таксонов

Динамическая типология

- При достаточно долгом существовании система стремится к состоянию динамического равновесия, при котором

$$\frac{N(A1)}{N(A2)} = \frac{p(A1 \Rightarrow A2)}{p(A2 \Rightarrow A1)}$$

- Вопрос: при каком значении $N(A2)$ система будет в равновесии?



Распределение «редких» (= ареально-специфических) черт

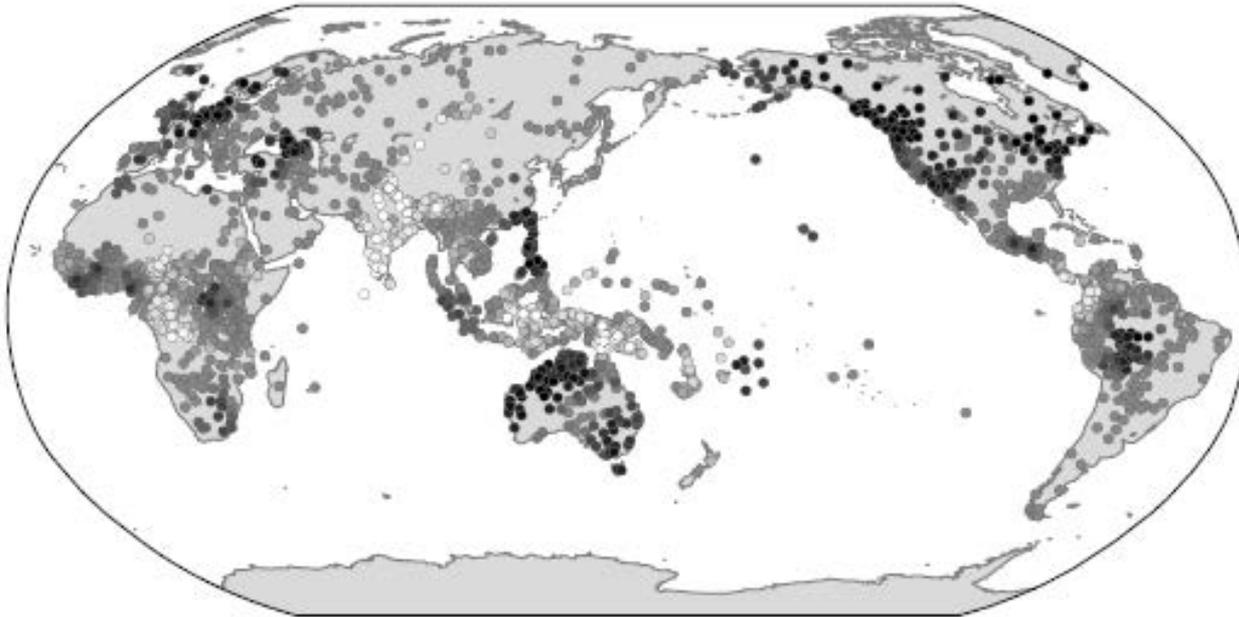


Figure 3. World map showing areal centres of rarity.

- Редкие черты распространены ареально неравномерно (например северозападная Европа, Кавказ и Филиппины)
- Объективный способ найти центры (не совсем традиционно понимаемых) ареалов.

Cysouw, Michael. 2011. Quantitative explorations of the world-wide distribution of rare characteristics, or: the exceptionality of northwestern European languages. In Horst Simon and Heike Wiese, editors, *Expecting the Unexpected*. Berlin: De Gruyter Mouton. 411–431.

Экологические объяснения

Высота над уровнем моря и абруптивы



Dark circles represent languages with ejectives, clear circles represent those without ejectives. Clusters of languages with ejectives are highlighted with white rectangles.

Спасибо!