

# Создание грамматики искусственного языка и программа порождения его словаря

И. Левин, А. Максимова  
руководитель Б. В. Орехов

Цель работы — создание такого языка, у которого типологически редкие, но возможные явления проявляют себя на всех уровнях.

Язык геикваку (yeikwak'u 'язык')

Для создания типологически редкой языковой системы были использованы типологические данные, в основном WALS. В качестве языковых черт брались самые редкие черты, в выборках по соответствующим переменным. Не брались близкие к уникальным и противоречащие друг другу языковые черты.

## Некоторые черты фонологии:

- только открытые слоги (13% языков)\*;
- ударение на третьем слоге с конца (2% языков)\*;
- 34 согласных;
- противопоставление по глухости/звонкости только у щелевых согласных (7% языков)\*;
- противопоставление придыхательных и эйективных взрывных согласных;
- 4 гласных ([i], [e], [a], [u]) и назализованные пары к ним.

\* Данные приводятся по WALS

## Некоторые черты грамматики:

- порядок слов OVS (0,8% языков)\*;
- косвенное дополнение между прямым дополнением и глаголом;
- преобладание двойного маркирования (7% языков)\*;
- преобладание префиксальной морфологии (6% языков)\*;
- эргативный строй;
- 5 падежей (маркируются препозиционными клитиками);
- будущее и небудущее время;
- маркированная неопределённость (аффикс на имени).

\* Данные приводятся по WALS

Словарь должен быть правдоподобным (похожим на словари естественных языков) и систематичным.

На правдоподобность влияет наличие сходств между лексемами.

Сходства между лексемами могут выражаться в

- наличии общего аффикса ('-ate' и '-er' у глаголов в английском и французском);
- наличии схожих морфологических изменений (gu-ranka 'нога', ma-tanka 'ноги' в языке фула);
- одинаковом словоизменении (части речи и именные классы).

Программа написана на Python 2.7 на базе словаря русского языка (~120 тыс. лексем).

Порождение случайной последовательности символов с рядом ограничений:

- от 1 до 4 слогов (более частотные лексемы короче);
- слог (C)V и фонемы с доп. артикуляцией реже других.

4 именных класса (маркеры '1', '2', '3', '4' в словаре): "Человек", "Животные", два класса без семантической мотивации (неравномерное случайное распределение).

Стриженная овца увидела коней, везущих тяжелую повозку с большим грузом, быстро несущих человека. (по А. Шлейхеру)

[[[тяжёлый=MOD INDEF-повозка с LOC=[большой=MOD INDEF-груз]] 3.NFUT-везти]=MOD [INDEF-человек быстрый 3.NFUT-нести]=MOD PL-INDEF-конь] 3.NFUT-PERF-видеть ERG=[3.NFUT-PERF-стричь=MOD INDEF-овца]

ts'ats'ēeu=t'unji ju-p'usi t'ats'e k'h'a-hajimu(u=t'unji ju-xip'ut's'i x'u-həvauk'uuyu=t'unji ju-k'e p'ets'hāi x'u-ŋəaxaet'e=t'unji ju-wi-juk'watu x'u-t'a-ŋāip'hēf'a ja=x'u-t'a-huk'hap'heni=t'unji ju-k'eeŋ'eŋa

## Перспективы:

- возможность применения языка для проработки вселенной какого-либо художественного произведения;
- создание автоматического переводчика.