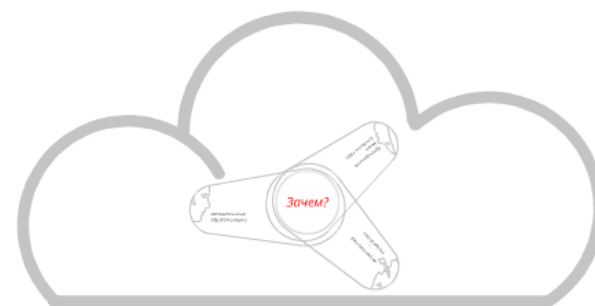


К построению частотной грамматики русского языка

Mihail.Kopotev@Helsinki.fi

В отличие от частотных словарей, где единицей является лексема (лемма), частотные грамматики имеют дело с разным единицами.

Частотная грамматика не термин, а "зонтичная" метафора, объединяющая представление грамматики языка на основе частотных характеристик грамматических единиц.



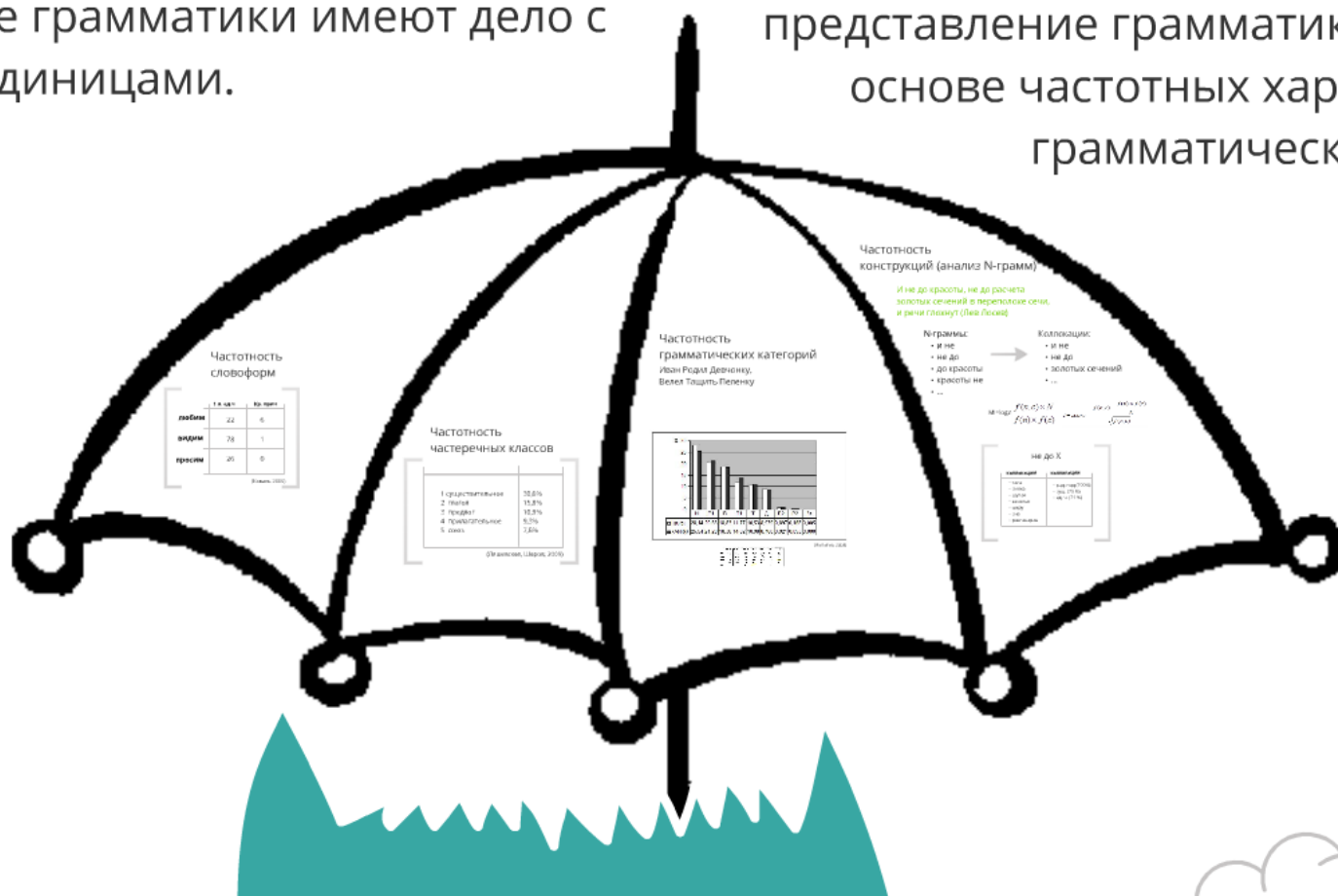
К построению частотной грамматики русского языка

Mihail.Kopotev@Helsinki.fi



В отличие от частотных словарей, где единицей является лексема (лемма), частотные грамматики имеют дело с разным единицами.

Частотная грамматика не термин, а "зонтичная" метафора, объединяющая представление грамматики языка на основе частотных характеристик грамматических единиц.



Частотность словоформ

	1 л. ед.ч	Кр. прич
любим	22	6
видим	78	1
просим	26	0

(Коваль 2006)

Частотность частеречных классов

1 существительное	30,6%
2 глагол	15,8%
3 предлог	10,9%
4 прилагательное	9,3%
5 союз	7,6%

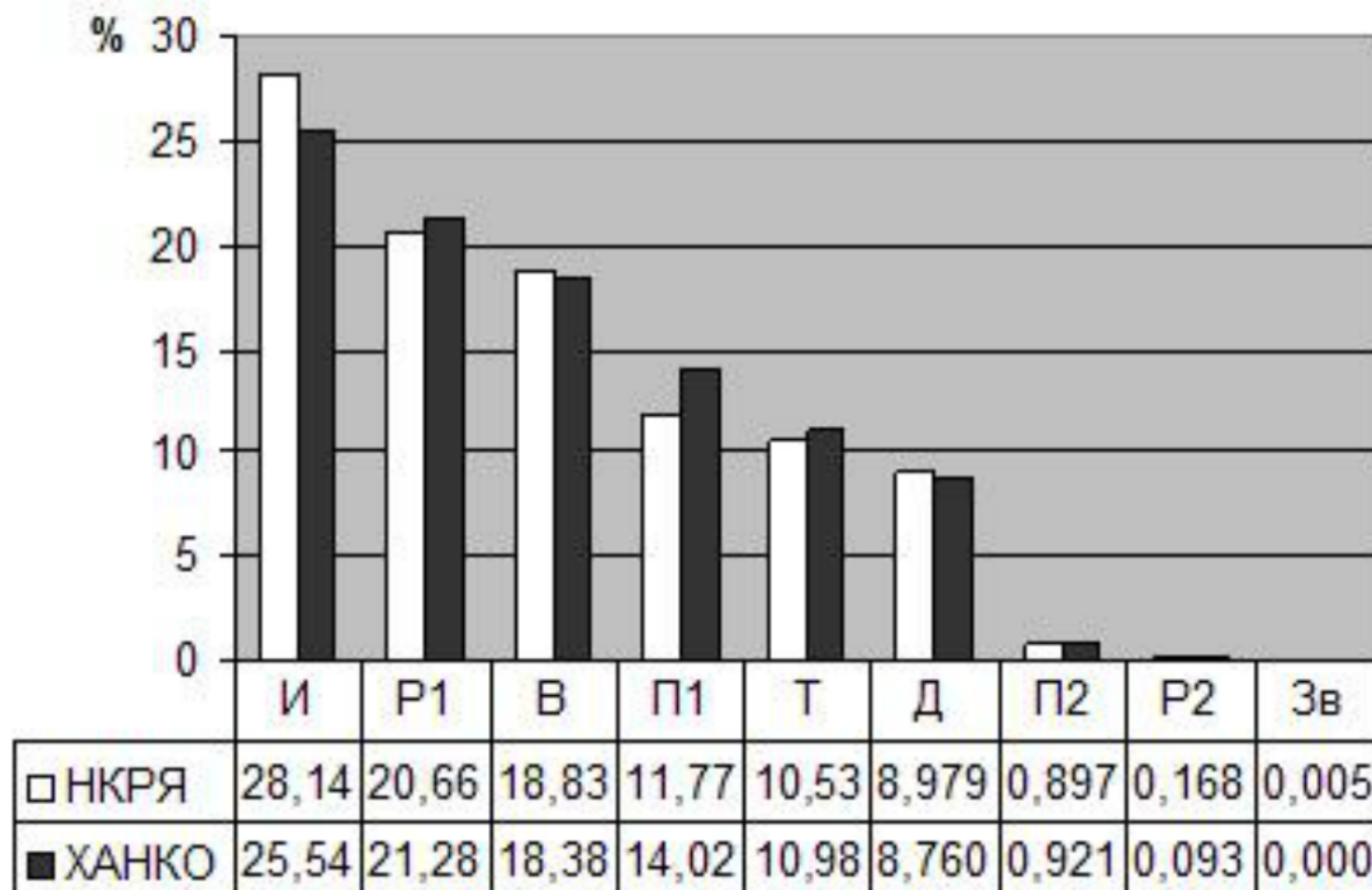
(Ляшевская, Шаров; 2009)

Частотность грамматических категорий

Иван Родил Девчонку,
Велел Тащить Пеленку

% 30





(Копотев, 2008)

	Ощ, %	Преп, %	Прет, %	SPPO, %	APPO, %	Числ, %	Числ преп, %
■ И	23,12	24,21	29,14	42,19	27,43	32,77	23,54
□ Р1	19,84	19,12	18,04	10,00	15,21	10,57	18,17
□ В1	16,70	17,02	18,20	14,57	21,45	44,42	22,60
■ П1	15,41	14,29	10,05	6,61	16,03	4,39	19,15
□ Т	14,84	19,12	17,57	8,18	12,25	5,20	12,52
□ Д	8,63	6,24	6,99	18,46	7,62	2,66	4,02

	Сущ., %	Прип., %	Прич., %	SPRO, %	APRO, %	Числ., %	Числ-прип., %
■ И	23,12	24,21	29,14	42,19	27,43	32,77	23,54
□ Р1	19,84	19,12	18,04	10,00	15,21	10,57	18,17
■ В1	16,70	17,02	18,20	14,57	21,45	44,42	22,60
■ П1	15,41	14,29	10,05	6,61	16,03	4,39	19,15
■ Т	14,84	19,12	17,57	8,18	12,25	5,20	12,52
■ Д	8,63	6,24	6,99	18,46	7,62	2,66	4,02

Частотность конструкций (анализ N-грамм)

И не до красоты, не до расчета
золотых сечений в переполохе сечи,
и речи глохнут (Лев Лосев)

N-граммы:

- и не
- не до
- до красоты
- красоты не
- ...



Коллокации:

- и не
- не до
- золотых сечений
- ...

$$MI = \log_2 \frac{f(n, c) \times N}{f(n) \times f(c)} \quad t\text{-score} = \frac{f(n, c) - \frac{f(n) \times f(c)}{N}}{\sqrt{f(n, c)}}$$

не до X

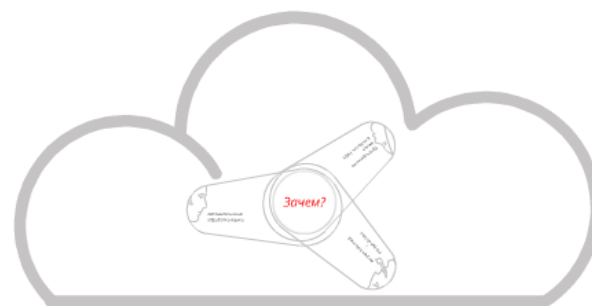
коллокации	коллигации
• того • смеха • шуток • веселья • жиру • сна • разговоров	• род. пад (100%) • сущ. (73 %) • ед. ч. (71 %)

К построению частотной грамматики русского языка

Mihail.Kopotev@Helsinki.fi

В отличие от частотных словарей, где единицей является лексема (лемма), частотные грамматики имеют дело с разным единицами.

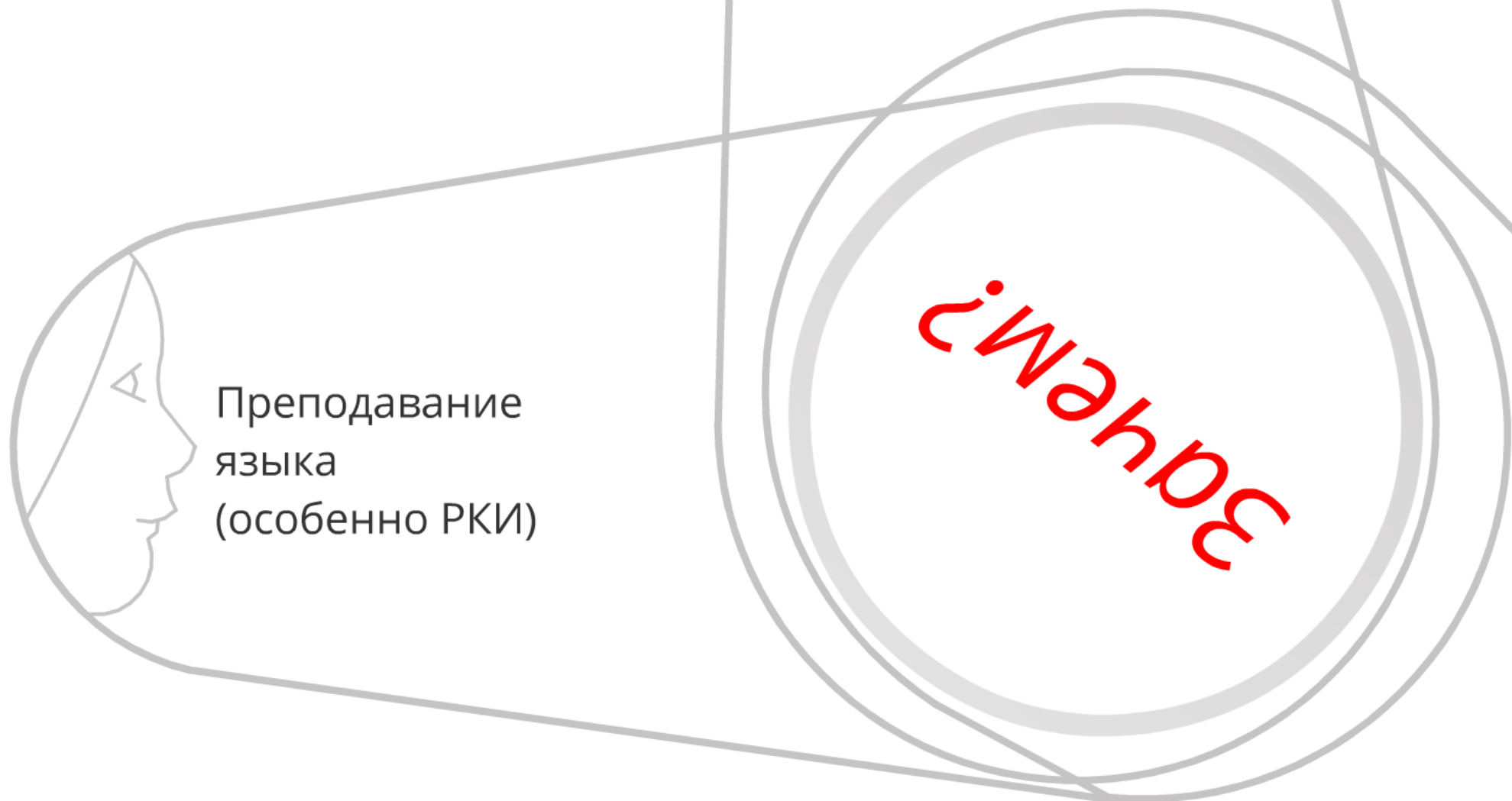
Частотная грамматика не термин, а "зонтичная" метафора, объединяющая представление грамматики языка на основе частотных характеристик грамматических единиц.





Зачем?

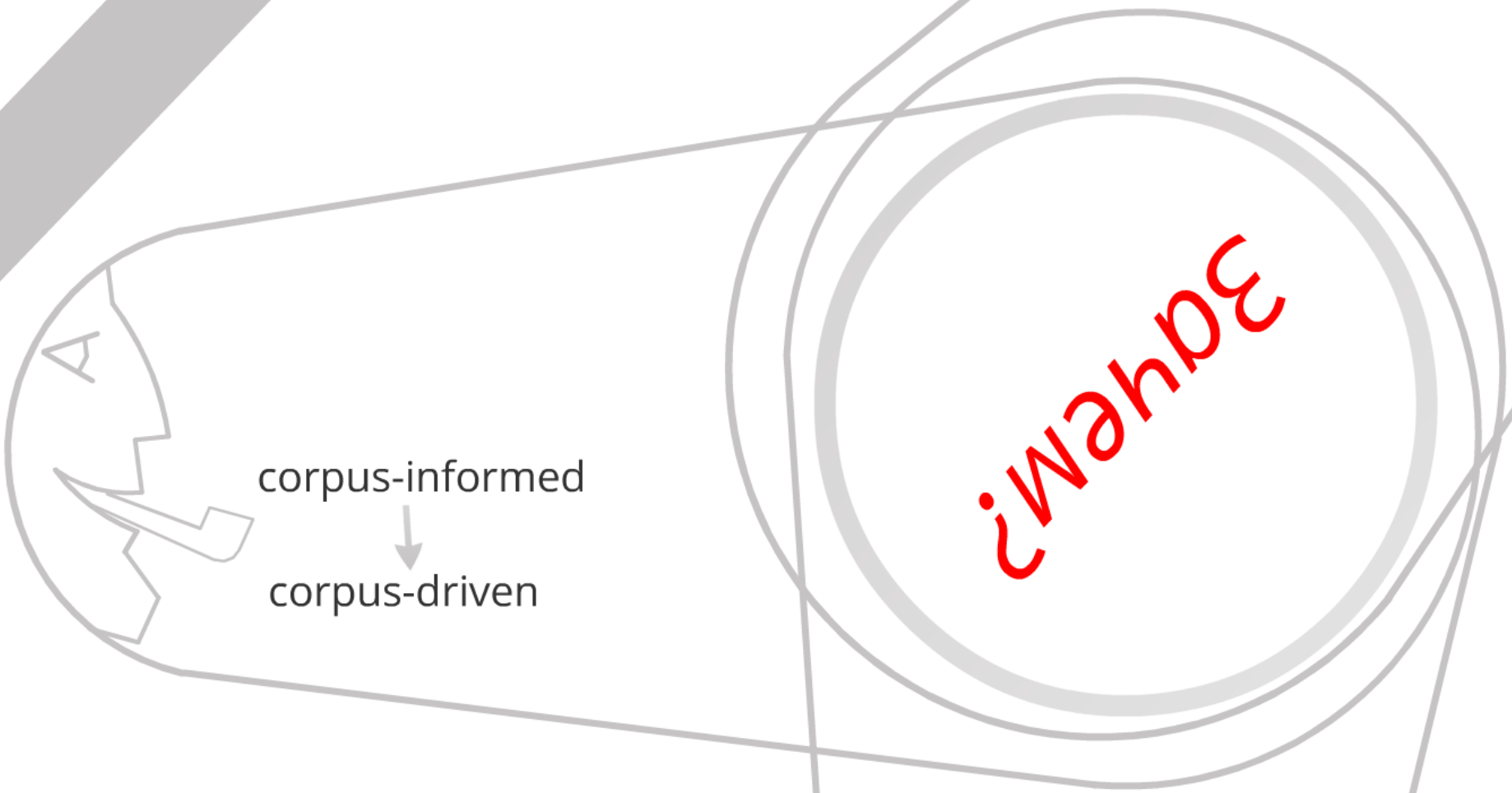
Автоматическая
обработка языка



Преподавание
языка
(особенно РКИ)

Зачем?

Тическая
языка



ание

РКИ)

К построению частотной грамматики русского языка

Mihail.Kopotev@Helsinki.fi

В отличие от частотных словарей, где единицей является лексема (лемма), частотные грамматики имеют дело с разным единицами.

Частотная грамматика не термин, а "зонтичная" метафора, объединяющая представление грамматики языка на основе частотных характеристик грамматических единиц.

