

Норма, кодификация и их верификация (демонстрация эффективности сегментно-статистического метода)

В. И. Беликов, ИРЯ РАН

На диске (но не в Сборнике материалов) есть большой материал из четырех частей:

1. Методические соображения о лексикографически ориентированном поиске в Интернете — (0,2 а. л.);
2. «Непобедимая нога»: спортивная лексика в нормативном словаре [Комплексный нормативный словарь, СПб, дек. 2009] — (0,9 а. л.);
3. О частоте встречаемости слов «в Интернете» — (0,2 а. л.);
4. О словарях, «содержащих нормы современного русского литературного языка при его использовании в качестве государственного языка Российской Федерации» — (4,3 а. л.).

В. А. Ицкович (Очерки синтаксической нормы, 1982):

Норма это «комплекс закрепленных речевой практикой языковых средств и закономерностей их реализации», объективно существующих «в данное время в данном языковом коллективе». «Имплицитно норма выступает в виде образца или, точнее, текстов, считааемых образцовыми <...> Эксплицитно, в явном виде, сформулированной, норма предстает перед носителями языка в кодификации <...> **Кодификация** — это фиксация объективно существующей языковой нормы, сформулированная в виде правил (предписаний) в авторитетном лингвистическом издании».

Литературной (в т. ч. разговорной) нормой удобно считать то, что служит образцом для большинства носителей литературного языка «в данное время в данном языковом коллективе».

Синхронно норма имеет **возрастное** и **территориальное** измерение.

Кодификация лексики — не только (субъективное) мнение лексикографа об объективной норме. Значительная часть словаря всегда получала разного уровня официальное утверждение.

«Российская газета» пишет о входных дверях московского метро:

*Их так мотает от ветра,
что и глазом не моргнешь,
как получишь по лбу (22.10.2009).*

Писать так можно в частных письмах и романах, а в газетах 1.09.09 — нельзя. (К счастью, для нарушителей закона пока не введены санкции)

Способы верификации любых языковых фактов:
интроспекция, обработка текстов
наблюдение, • «традиционных»
работа с информантом. • оцифрованных

Общепринятые процедуры верификации фактов, касающихся лексики, на материале оцифрованных текстов:

- «Гугление», в том числе Яндексом, «в Интернете вообще»:

дает **идеальные** результаты для **исключительно редких** словоупотреблений при аккуратной (!) работе.

- Поиск в НКРЯ:

дает **хорошие** результаты для **частотных** словоупотреблений.

«Интернет вообще», содержательная сторона:

Один коллега провел небольшое исследование «**Чьи стихи нынче любят?**». Искались текстовые фрагменты типа "*люблю стихи X-а*".

В выдачу попали тексты:

- ***Я люблю стихи Коржавина, ценю его публицистику.*** (Довлатов, «Синтаксис», 1982). — почти нынче.
- ***...так нежно я люблю стихи Грибоедова и его самого...*** (К. С. Станиславский)
- ***Вы любите стихи, Антон Павлович? — Да, я очень люблю стихи Пушкина*** (Л. А. Авилова. А. П. Чехов в моей жизни).

Кроме героя незавершенного романа Станиславского, стихи Грибоедова любил только один блоггер 2 февраля 2009 г.

Для осмысленности результатов поиска Интернет нужно фрагментировать на относительно однородные сегменты.

«Соборное уложение» Алексея Михайловича и блог 12-летней поклонницы японских аниме написаны на разных языках.

Типичное утверждение любителя гугления:

Например, написание *Таллинн* встречается в Интернете 6 млн раз, а *Таллин* 4 млн раз, употребление предложно-падежной словоформы *в Украине* использовано на 62 млн страниц Интернета, а словоформа *на Украине* употреблена на 60 млн страниц. Написание *Кыргызстан* использовано 6 млн, а *Киргизия* — 10 млн раз.

С. А. Кузнецов, Языковая норма и правила речевой деятельности // Комментарий к Федеральному закону «О государственном языке Российской Федерации». Часть 1. Доктринальный и нормативно-правовой комментарий, стр. 32.

При запросе "Украину" должны находиться тексты типа Люблю *Украину*, инвестировать *в Украину*, волнуюсь *за Украину*, еду *на Украину*, подкоп *под Украину* через Тузлу и прочее.

Всего "Украину" «нашлось на 136 млн страниц»:

"Украину" — Яндекс: нашлось 136 млн страниц - Windows Internet E...

Я http://yandex.ru/yandsea Яндекс

Я "..." Я Расп...

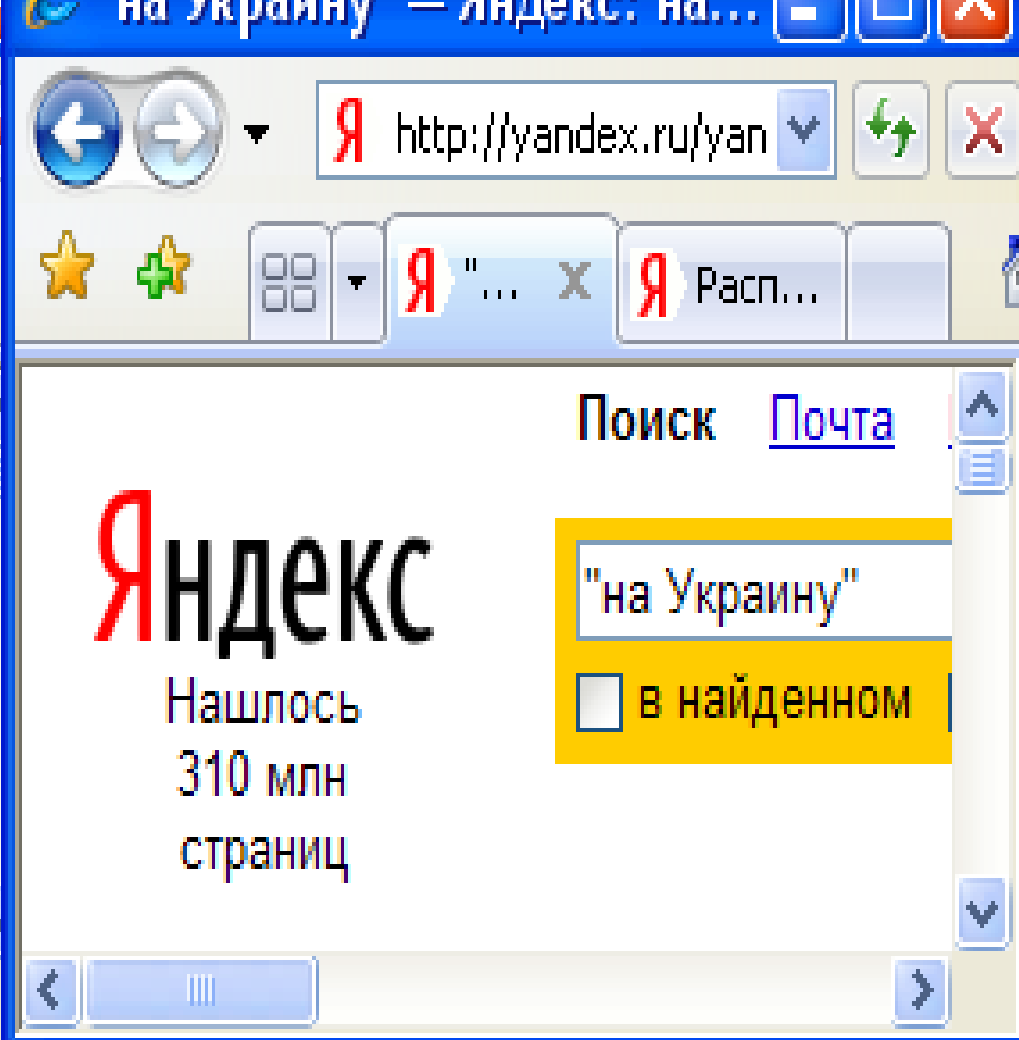
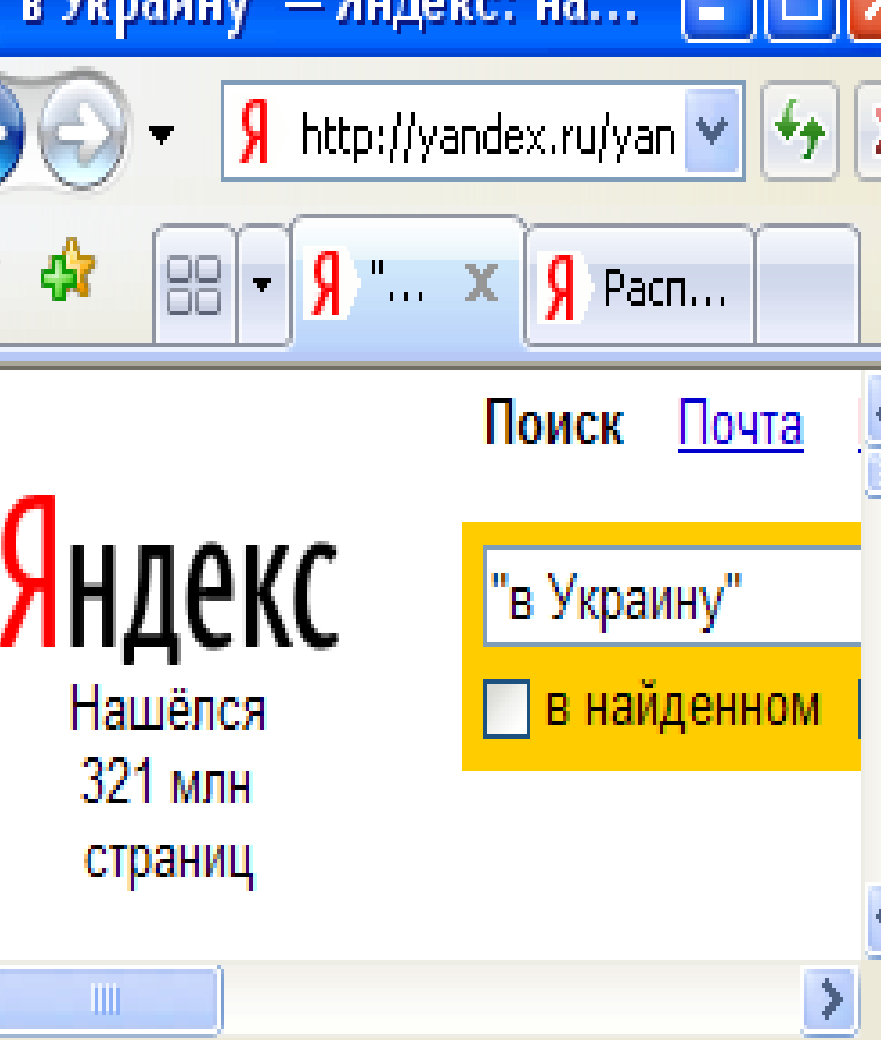
Поиск [Почта](#) [Карты](#) [Маркет](#) [Новости](#) [Слов...](#)

ЯНДЕКС

Нашлось
136 млн
страниц

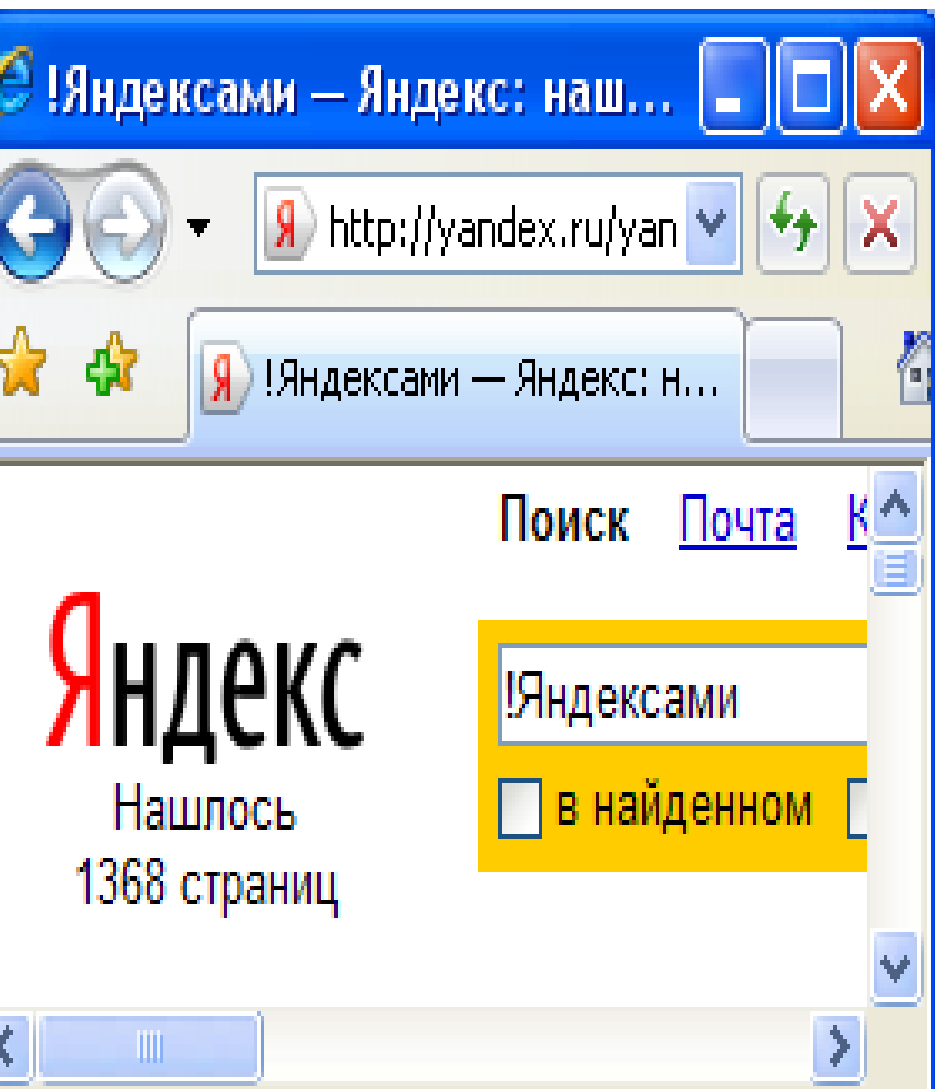
"Украину"

в найденном в Москве



Любители высказывания «... встречается в Интернете ... млн раз» легко решат уравнение: $321 + 310 + x = 136$.
(я не решал — жалко времени на интерпретацию)

Поиск словоформы: *"Яндекссами"* vs. *!Яндекссами*.
Результаты различаются в два раза и более:



!Яндекссами — Яндекс: наш...

http://yandex.ru/yan

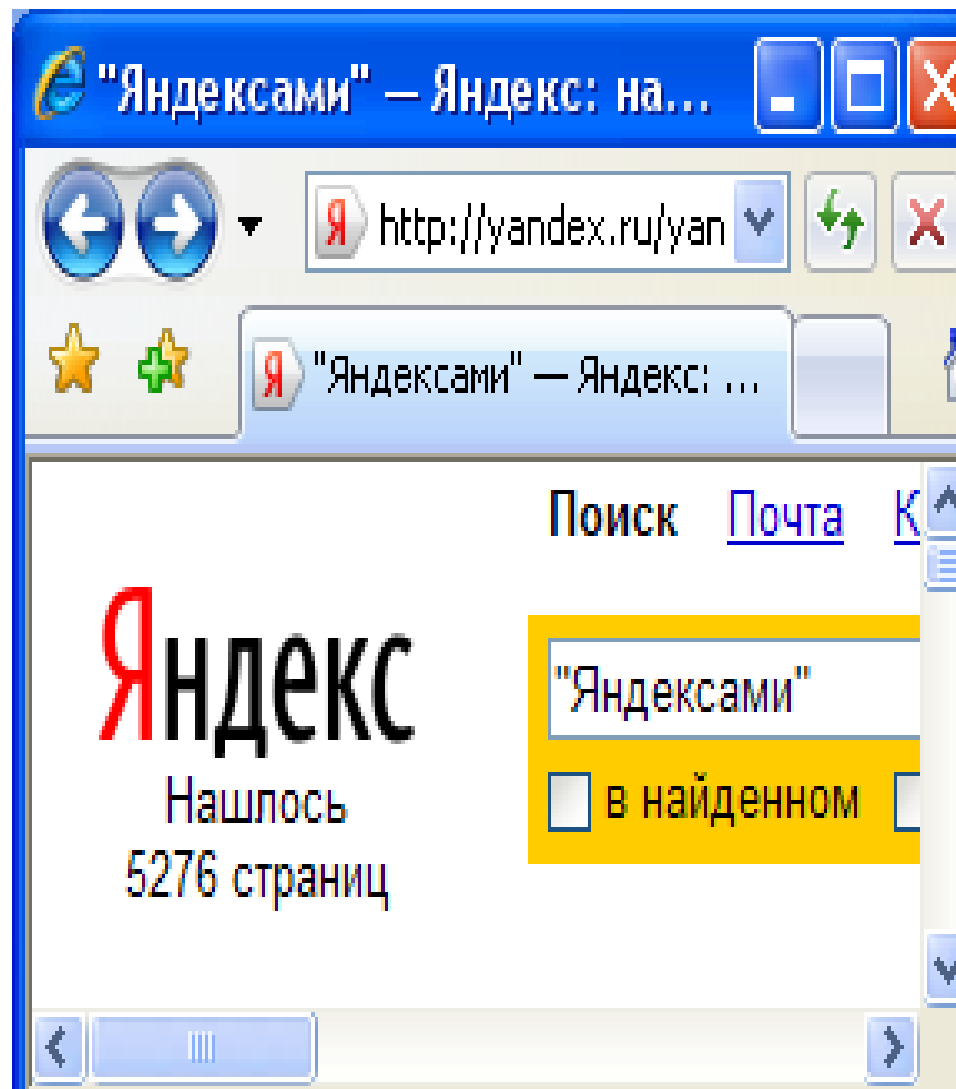
!Яндекссами — Яндекс: н...

Поиск [Почта](#) [К](#)

Яндекс

Нашлось
1368 страниц

в найденном



"Яндекссами" — Яндекс: на...

http://yandex.ru/yan

"Яндекссами" — Яндекс: ...

Поиск [Почта](#) [К](#)

Яндекс

Нашлось
5276 страниц

в найденном

«Ну, подумаешь, Яндекс...
Вот Google...»

Таллинн погугленный в деталях таков:

Таллинн без кавычек: **2 млн 40 тыс.** страниц.

Поиск по отдельным словоформам:

<i>Таллинн</i>	192000	<i>Таллинны</i>	785
<i>Таллинна</i>	264000	<i>Таллиннов</i>	71
<i>Таллинну</i>	146000	<i>Таллиннам</i>	36
<i>Таллинном</i>	98700	<i>Таллиннах</i>	39
<i>Таллинне</i>	396000	<i>Таллиннами</i>	34

всего ед. ч.: 1096700

всего мн. ч. 965 ед.+мн.: **1097665**

Итого: совокупно **на другие словоформы ед. и мн. числа**,
а также на словоформы двойственного и других чисел,
не омонимичные просмотренным:

$2040000 - 1097665 = \mathbf{942335}$

(скриншоты для поклонников Google'a — на компьютере)

НКРЯ vs. разные сегменты Интернета как выразить по-русски Oper ([кассовый] чек)?

1. субъект — кассир/продавец или кассовый аппарат,
2. субъект — покупатель.

- **Словари** противоречат друг другу
- **В узусе:** *выби(ва)ть / проб(ва)ть* с неясным распределением а также *отби(ва)ть* чек, которому многие изумляются.
- **НКРЯ:**
выбить/пробить/отбить на расстоянии от 1 до 10 от чек:
По этому запросу ничего не найдено.
выбивать/пробивать/отбивать: **19** (13/5/1)
(11 — с глаголами сов. вида)
- «**Журнальный зал**» (magazines.russ.ru) : **41** (19/15/7)
Примеров достаточно для вполне обоснованных гипотез, которые полностью подтверждаются на других сегментах.
(Все детали демонстрируются на компьютере)

Спасибо за внимание!

Дополнительное Большое Спасибо

тем, кто проявит любопытство
и подойдет ко мне

во время основной части стендовой сессии