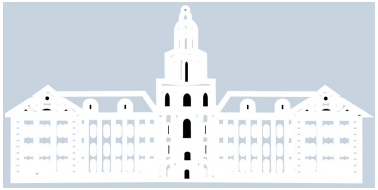

Лекция 3

# Разметка (мультимедийного) корпуса

Ольга Ляшевская ** olesar@yandex.ru

Курс “Лингвистические данные”, 1 курс ФиКЛ ВШЭ

# В прошлой лекции

- Корпус: слои разметки, соответствующие основным уровням описания
  - леммы
  - грамматические разборы форм словоизменения
  - морфемно-словообразовательные разборы
  - лексико-семантические группы
  - синтаксическая структура
  - семантические роли, кореференция
  - фонетика
  - интонация
  - жесты
- Переводы, глоссы
- Специальная разметка (стиховедческая, в параллельных корпусах и т.д.)
- Метаданные

- Корпус: слои разметки, соответствующие основным уровням описания
  - леммы
  - грамматические разборы форм словоизменения
  - морфемно-словообразовательные разборы
  - лексико-семантические группы
  - синтаксическая структура
  - семантические роли, кореференция
  - фонетика
  - интонация
  - жесты
- Переводы, глоссы
- Специальная разметка (стиховедческая, в параллельных корпусах и т.д.)
- Метаданные

# Разнообразие инвентарей

- Грамматические разборы, примеры:
  - LOB corpus all_ABN the_ATI girls_NNS love_VB a_AT scholar_NNS ._

    CLAWS1 tagset

# Разнообразие инвентарей

- Грамматические разборы, примеры:
  - LOB corpus all_ABN the_ATI girls_NNS love_VB a_AT scholar_NNS ._
  - MULTEXT (Sketch Engine)

**Vmip3p-m-e- (RUS: *курятся*)**

| Category | Code | Attributes | Values | Languages |
|---|---|---|---|---|
| Noun | N | 14 | 68 | 16 |
| Verb | V | 17 | 74 | 16 |
| Adjective | A | 17 | 79 | 16 |
| Pronoun | P | 19 | 97 | 16 |
| Determiner | D | 10 | 32 | 3 |
| Article | T | 6 | 23 | 3 |
| Adverb | R | 7 | 28 | 16 |
| Adposition | S | 4 | 12 | 16 |
| Conjunction | C | 7 | 21 | 16 |
| Numeral | M | 13 | 81 | 16 |
| Particle | Q | 3 | 17 | 12 |
| Interjection | I | 2 | 4 | 16 |
| Abbreviation | Y | 5 | 35 | 16 |
| Residual | X | 1 | 3 | 16 |

| | | |
|---|---|---|
| V | Category | Verb |
| m | Type | main |
| i | VForm | indicative |
| p | Tense | present |
| 3 | Person | third |
| p | Number | plural |
| - | Gender | - |
| m | Voice | media |
| - | Definiteness | - |
| e | Aspect | perfective |
| - | Case | - |

# Разнообразие инвентарей

- Грамматические разборы, примеры:
  - Universal dependencies tagset vs. Stockholm Umeå Corpus tagset

| | | | | | |
|---|---|---|---|---|---|
| 1 | Då | då | ADV | AB | _ |
| 2 | var | vara | VERB | VB.PRET.ACT | Tense=Past\|Voice=Act |
| 3 | han | han | PRON | PN.UTR.SIN.DEF.NOM | Case=Nom\|Definite=Def\|Gender=Com\|Number=Sing |
| 4 | elva | elva | NUM | RG.NOM | Case=Nom\|NumType=Card |
| 5 | år | år | NOUN | NN.NEU.PLU.IND.NOM | Case=Nom\|Definite=Ind\|Gender=Neut\|Number=Plur |
| 6 | . | . | PUNCT | DL.MAD | _ |

# Промежуточные выводы

- Грамматическая разметка, инвентарь помет (а также и другие виды разметки) зависят от свойств языка, представленных в корпусе вариантов, теоретических традиций описания языка, задач пользователей

- Тем не менее, корпусная лингвистика успешно решает задачу приравнивания и конверсии одних систем помет в другие

# Разметка невербальной коммуникации

- Голосовой канал - интонация фразы, слова, фокуса
- Тело - жест, мимика, движения глаз, окулометрия, мышечные сокращения и другая физиология

# Литература


- Гришина Е. А. Русская жестикуляция с лингвистической точки зрения. Корпусные исследования. М., 2017.
- Litvinenko A. O., Kibrik A. A., Fedorova O. V., Nikolaeva J. Annotating hand movements in multichannel discourse: Gestures, adaptors and manual postures // Российский журнал когнитивной науки 5(2), 2018. Pp. 4–17.
- Кибрик А. А. 2018. Русский мультиканальный дискурс. Часть I. Постановка проблемы // Психологический журнал 39(1). 70–80.
- Кибрик А. А. 2018. Русский мультиканальный дискурс. Часть II. Разработка корпуса и направления исследований // Психологический журнал 39(2). 79–90.