

Зачем нужен правильный морфологический анализ в XXI веке?

Язык, текст, культура через призму цифровых технологий

НИУ ВШЭ СПб

04–05 декабря 2025

Г. А. Мороз НИУ ВШЭ, Москва

Международная лаборатория языковой конвергенции

5.12.2025

Зачем нужен морфологический анализ?

Что такое морфологический анализ

Морфологический анализ, как его обычно видят лингвисты, обычно включает в себя несколько вещей:

- определение морфологической формы (например, латинское *mensam* — acc.sg);
- приведение к начальной форме или основе (*mensam* — *mensa*, ж. р., первое склонение);
- перевод основы (*mensam* — ‘стол’).

Что такое морфологический анализ

Морфологический анализ, как его обычно видят лингвисты, обычно включает в себя несколько вещей:

- определение морфологической формы (например, латинское *mensam* — acc.sg);
- приведение к начальной форме или основе (*mensam* — *mensa*, ж. р., первое склонение);
- перевод основы (*mensam* — ‘стол’).

Не бывает правильного морфологического анализа — это каждый раз компромисс между языковой моделью, конвенциями исследователей и текущими задачами. Например, лингвисты, приводя примеры, редко перечисляют несловоизменительную информацию (например, род для существительных).

Зачем нужен морфологический анализ?

- Польза от морфологически размеченных текстов для теоретической лингвистики неоценима

Зачем нужен морфологический анализ?

- Польза от морфологически размеченных текстов для теоретической лингвистики неоценима
- В области NLP — все не так однозначно:
 - Какие-то задачи можно решить при помощи *стэминга*
 - Выгода зависит от языка к языку
 - Легко представить, что в некоторых задачах выгоднее иметь лишь какую-то часть морфологического анализа, например, частеречную разметку

Подходы к морфологическому анализу

Много данных

Обычно, если много данных, люди используют нейросети. Для морфологического анализа русского языка их использовали в следующих работах [Arefyev et al., 2018, Sorokin and Kravtsova, 2018, Bolshakova and Sapin, 2019a,b, 2020, Garipov et al., 2023]. Используются разные архитектуры:

- свёрточные нейронные сеть (convolutional neural network, CNN);
- деревья решения с градиентным бустингом (decision trees with gradient boosting);
- двунаправленная длинная цепь элементов краткосрочной памяти (Bidirectional long short-term memory network, Bi-LSTM);
- и другие.

hi

Морфологические трансдьюсеры

Где почитать еще?

- **Онлайн материалы** нашего с Т. Казаковой курса в НИУ ВШЭ
- **ACL Anthology:**
 - morphological transducers — около 3440 результатов
 - transducers — около 6450 результатов

Спасибо за внимание!

Ссылки на литературу I

- N. V. Arefyev, T. Y. Gratsianova, and K. P. Popov. Morphological segmentation with sequence to sequence neural network. In *Компьютерная лингвистика и интеллектуальные технологии*, pages 85–95, 2018.
- E. I. Bolshakova and A. S. Sapin. Bi-lstm model for morpheme segmentation of russian words. In *Artificial Intelligence and Natural Language: 8th Conference, AINL 2019, Tartu, Estonia, November 20–22, 2019, Proceedings* 8, pages 151–160. Springer, 2019a.
- E. I. Bolshakova and A. S. Sapin. Comparing models of morpheme analysis for russian words based on machine learning. In *Компьютерная лингвистика и интеллектуальные технологии*, pages 104–113, 2019b.

Ссылки на литературу II

- E. I. Bolshakova and A. S. Sapin. An experimental study of neural morpheme segmentation models for russian word forms. In *CMCL*, pages 79–89, 2020.
- T. Garipov, D. Morozov, and A. Glazkova. Generalization ability of cnn-based morpheme segmentation. In *2023 Ivannikov Ispras Open Conference (ISPRAS)*, pages 58–62. IEEE, 2023.
- Alexey Sorokin and Anastasia Kravtsova. Deep convolutional networks for supervised morpheme segmentation of russian language. In *Artificial Intelligence and Natural Language: 7th International Conference, AINL 2018, St. Petersburg, Russia, October 17–19, 2018, Proceedings 7*, pages 3–10. Springer, 2018.