



МГУ им. М.В.Ломоносова  
Научно-исследовательский  
вычислительный центр



АНО Центр  
информационных  
исследований

---

---

Лукашевич Н.В

**Квазисинонимы в лингвистических  
онтологиях**

***Near-synonyms in linguistic ontologies***

# Онтологии для автоматической обработки текстов

- **Основные элементы онтологий**
  - понятия онтологии,
  - отношения между понятиями
- **Автоматическая обработка текстов**
  - Необходимо установление отношений между понятиями онтологии и языковыми выражениями
- **Лингвистические (лексические) онтологии – онтологии, которые учитывают систему значений естественного языка**

# Лингвистические онтологии и проблемы лексической семантики

- Каждое понятие лингвистической онтологии имеет совокупность текстовых выражений, которое может это понятие выразить
- Совокупность текстовых выражений одного понятия – **онтологические синонимы**
- Проблемы описания
  - Близкие значения слов нужно дискретно подразделить,
  - **Квазисинонимы нужно разбить на совокупность дискретных синонимических рядов – выделить онтологические синонимы**

# План презентации

- **Квазисинонимы и проблемы отражения квазисинонимов в лингвистических онтологиях**
- **Пример из WordNet: квазисинонимы со значением =сходство=**
- **Принципы рассмотрения квазисинонимов в тезаурусе RuТез**
- **Примеры анализа квазисинонимов**

# Представление квазисинонимов в лингвистических онтологиях

- **Квазисинонимы (частичные синонимы)**
  - Слова с близкими значениями
  - Могут быть взаимозаменяемыми в некоторых контекстах
- **Квазисинонимы могут различаться по многим параметрам:**
  - Денотативный статус, стилистическая окраска,
  - Оценка, Коллокации, Валентности и др.
- **В разных контекстах могут употребляться в более широком или более узком смысле**
- **Близкий ряд квазисинонимов на другом языке характеризуется своей собственно системой различий**

# Примеры квазисинонимов

- ошибка, погрешность, недосмотр, просмотр, ляп, промах, оплошность, осечка, прокол, упущение, недочет, а также ослышка, описка, опечатка, оговорка.
- error, fault, omission, oversight, blunder, mistake, miss, screw-up, dereliction, defect
- Драться, поддаться, переддаться, свалка, потасовка, побоище, мордобой, поножовщина

# Квазисинонимы в предметных областях

- *кредитование, кредит, кредитная услуга,*
- *кредитное обслуживание, кредитная операция,*
- *выделение кредита, выдача кредита,*
- *выделение кредитных средств,*
- *предоставление кредита*
- 
- *Каковы особенности отображения квазисинонимов в онтологии?*

# Рекомендации по созданию онтологий: различимость понятий

- Нужно различать понятия и его названия: не стоит заводить отдельные понятия для синонимов
- Понятие-потомок должно отчетливо отличаться от понятия-родителя
- Понятие должно быть отчетливо отличаться от понятий одного уровня
- Это важно:
  - для описания отношений;
  - для формального вывода;
  - для перевод на другой язык

# Квазисинонимы in WordNet

- Основная единица – синсет
  - Совокупность синонимов
- 4 плохо отличимых синсета, описывающих сходство. Каждый синсет – гипоним предыдущего:
- *sameness* –(the quality of being alike)
- *similarity* – (the quality of being similar)
- *likeness, likeness, similitude* – (similarity in appearance or character or nature between persons or things)
- *resemblance* – (similarity in appearance or external or superficial details)

# Тезаурус РуТез - лингвистическая онтология



- ❖ **Понятие:**
  - ❖ **Имя понятия**
  - ❖ **Набор текстовых выражений**
  - ❖ **Отношения между понятиями**
- ❖ **52 000 понятий,  
156 000 текстовых выражений,  
203 000 отношений (более 2 млн. с иерархией)**
- ❖ **Переведен на английский язык:  
130 тысяч слов и выражений**
- ❖ **Приложения информационного поиска:  
формулировка запросов, автоматическое расширение  
запросов, автоматическая рубрикация, кластеризация,  
аннотирование**

# Понятия в тезаурусе RuТез: основные принципы

- **Различимые понятия**
  - **разный набор отношений с другими понятиями тезаурус,**
  - **разный набор онтологических синонимов**
- **Традиция информационно-поисковых тезаурусов - однозначное и понятное имя,**
- **Онтологические синонимы должны быть эквивалентны относительно системы отношений с другими понятиями тезауруса**

# Имя понятия: примеры

- - **однозначное слово:**
  - *КАБЕЛЬ;*
- - **однозначное словосочетание:**
  - *КАБИНЕТ РЕСТОРАНА,*
  - *КАБИНЕТ ВРАЧА*
- - **неоднозначное слово с пометой:**
  - *КАБАЧОК (ПЛОД);*
- - **пара синонимов – текстовых входов понятия через запятую:**
  - *ИРРАЦИОНАЛЬНЫЙ, ЛОГИЧЕСКИ НЕОБЪЯСНИМЫЙ*

# Словосочетания - синтаксические СИНОНИМЫ МНОГОЗНАЧНЫХ СЛОВ

- *авангард3 = авангардное искусство*
- *авангард4 = произведения авангарда*
- *чай3 = настой чая*
- *бородка2 = бородка ключа*
- *болид1 = космический болид*
- *болид2 = гоночный болид*
- *блок1 = подъемный блок*
- *экспедиция2 = отдел экспедиции*
- ...

# Основные принципы работы с квазисинонимами

- Искать различия между квазисинонимами, которые не исчезают в зависимости от контекста их употребления
- Искать различия между квазисинонимами, которые приводят к формированию разных рядов онтологических синонимов или к разным отношениям с другими понятиями
- Фиксировать найденные различия вводом понятий с однозначными именами

# Процедура ввода понятий для квазисинонимов (similarity)-1

- 0 шаг: ввод обобщенного понятия для квазисинонимов
  - SIMILARITY
- 1 шаг: найти признаки, по которым могут отличаться понятия
  - Сходство по внешнему виду - similarity in appearance
- 2 шаг: сформулировать имя понятия
  - Должно быть однозначным,
  - Лучше реально употребляющееся словосочетание
  - SIMILARITY IN APPEARANCE
    - 34700 страниц в GOOGLE

# Процедура ввода понятий для квазисинонимов-2

- Шаг 3. Найти разнообразные онтологические синонимы для этого понятия
  - *resemblance in appearance*,
  - *similarity of appearance*,
  - *external resemblance*
- Шаг 4. Многозначные слова, употребляемые в разных контекстах то в более общем смысле, то в более узком – поставить онтологическими синонимами к двум понятиям
  - *resemblance*
  - *likeness*

***SIMILARITY***  
resemblance, likeness

***SIMILARITY IN APPEARANCE***  
resemblance in appearance, similarity of appearance,  
external resemblance, resemblance, likeness, alikeness

***MUTUAL RESEMBLANCE***  
symmetrical resemblance

***SPLITTING IMAGE***

***MIRROR IMAGE***  
reflection, reflexion, mirror  
reflection, mirror symmetry,  
reflection symmetry

# Памятник, монумент (НОСС)

- - в память о конкретном человеке обычно ставится памятник, о группе людей – и памятник, и монумент, о событии – монумент; идеи воплощаются в монументах;
- - у монументов есть способность увековечивать подвиг живых людей;
- - по форме сооружения памятник часто представляет собой изображение увековечиваемого объекта;
- - монумент обычно больше по размерам;
- - пропагандистская роль больше свойственна монументам.
- **Онтологические синонимы или нужно заводить отдельные понятия?**

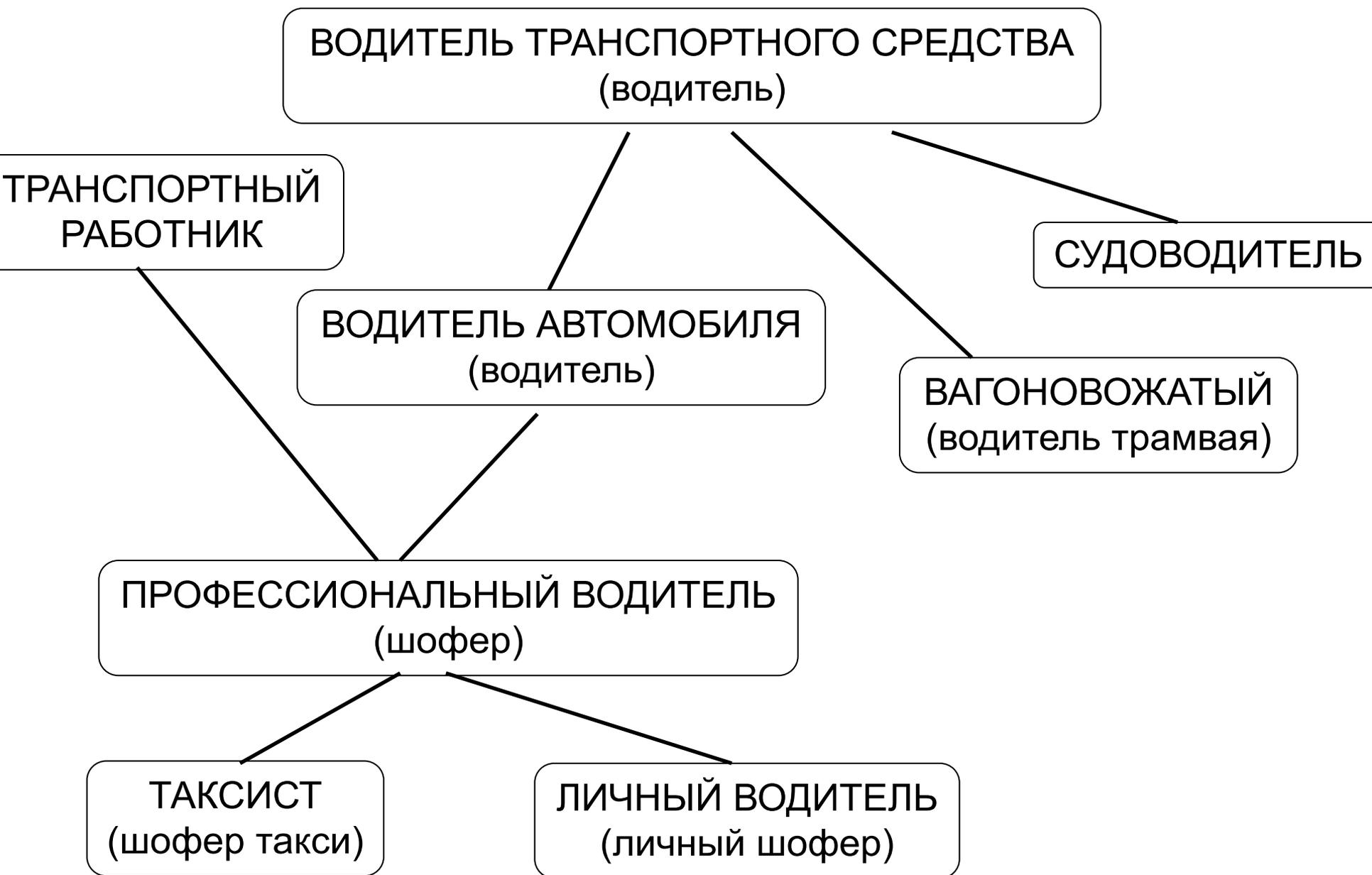
# Памятник, монумент-анализ примеров

- Указанные в словаре различия не являются обязательными
  - В память о конкретном человеке может быть установлен монумент;
  - В память события может быть установлен памятник;
  - Памятник может быть поставлен идее и т.п.
- Авторы словаря указывают, что различия «нейтрализуются при повторной, сокращенной номинации того же сооружения».
- Нет ни одного четко различающего свойства.
- **Памятник и монумент – онтологические синонимы**

# Водитель, шофер (НОСС)

- **НОСС: «шофер управляет только автомобилем или автобусом, водитель и другими транспортными средствами»**
- **Вагоновожатый, судоводитель являются водителями, но не шоферами**
- **Два понятия**
  - **ВОДИТЕЛЬ ТРАНСПОРТНОГО СРЕДСТВА,**
  - **ВОДИТЕЛЬ АВТОМОБИЛЯ**
- **Почему водитель и шофер ощущаются как синонимы?**

# Сеть понятий: водитель, шофер



# Заключение

- Важно стремиться создавать систему различных понятий даже для лингвистических онтологий
- Различимое понятие может отличаться набором онтологических синонимов и отношений с другими понятиями
- Важно формулирование однозначного, понятного имени понятия
- В этом помогает существование однозначных словосочетаний, синонимичным отдельным многозначным словам
- Если понятия отличимы, то сеть понятий может быть достаточно подробной
- Отличимые понятия делают онтологию менее зависимой от конкретного естественного языка