

Практическое занятие №3. Методы семантического анализа естественно-языковых текстов.

Цели работы

1. Ознакомиться с моделью «Смысл \leftrightarrow Текст».
2. Закрепить навыки работы с текстовыми данными.
3. Приобрести навыки работы с формальными лингвистическими моделями.

Задание

1. Выбрать из перечня лексем, взятых в предыдущих работах, 7 слов. Желательно, чтобы среди них были представлены разные части речи.
2. Для каждого из выбранных слов привести значения подходящих по смыслу лексических функций (см. краткие теоретические сведения и литературу)
3. Разработать формат хранения данных о значениях лексических функций слов (например, подмножество JSON или XML).
4. Реализовать 10 фрагментов кода на языке программирования Python, представленных в Pract_4_1.pdf и Pract_4_2.pdf

Краткие теоретические сведения

Модель «Смысл \leftrightarrow Текст» существенным образом использует большое количество сведений о каждом слове рассматриваемого языка (в нашем случае русского). Эти сведения могут быть представлены в особом словаре, который мы называем толково-комбинаторным словарем (ТКС) и который занимает в описываемой модели важное место. Словарная статья такого словаря должна в идеале содержать все словоизменительные, словообразовательные, синтаксические, семантические и стилистические характеристики заглавного, или ключевого, слова C_0 ; кроме того, для C_0 необходимо указать все слова (или словосочетания), определенным образом связанные с ним по смыслу, а именно: 1) его «парадигматические варианты», или «замены», — средства, которые могут или должны заменять C_0 в тех или иных контекстах и при тех или иных условиях; 2) его «синтагматические партнеры», или «параметры», — средства, которыми идиоматично, т. е. несвободно, выражаются при данном слове некоторые смыслы.

Для обозначения парадигматических вариантов и синтагматических партнеров слова мы будем использовать термин лексические корреляты. Зависимости, связывающие слова с их лексическими коррелятами, предлагается описывать с помощью лексических функций. А именно, лексическая функция (ЛФ) f описывает зависимость, определяющую для некоторого слова или словосочетания X такое множество слов или словосочетаний $\{Y^*\} = f(X)$, что для любых X_1 , X_2 верно следующее: если $f(X_1)$ и $f(X_2)$ существуют, то между $f(X_1)$ и X_1 с одной стороны, и между $f(X_2)$ и X_2 , с другой стороны, всегда имеет место одно и то же смысловое отношение. (Это можно изобразить символически, как ' $f(X_1)$ ': ' X_1 ' = ' $f(X_2)$ ': ' X_2 ')

X (ключевое слово или словосочетание) — это аргумент лексической функции f , а $\{Y_i\}$ — ее значение, или выражение. Лексические функции вводятся как средство описания лексической сочетаемости.

Среди ЛФ мы будем далее различать стандартные и нестандартные ЛФ. Стандартная в данном языке лексическая функция (СЛФ) должна — в отличие от нестандартной ЛФ — удовлетворять двум следующим требованиям:

(1) СЛФ имеет достаточно широкую семантическую сочетаемость, т. е. соответствующий смысл образует допустимые комбинации с достаточно большим числом разных смыслов. Иначе говоря, СЛФ должна быть определена для достаточно большого числа аргументов. Это свойство не зависит от языка.

(2) СЛФ характеризуется достаточным богатством ее языковых выражений. Это свойство, разумеется, полностью зависит от языка. Смысл 'имеющий большой вес' = 'тяжелый' обладает очень широкой сочетаемостью, т. е. удовлетворяет требованию (1); про огромное количество материальных предметов в принципе можно говорить, тяжелые они или не тяжелые. Однако в русском языке смысл 'тяжелый*' не удовлетворяет требованию (2), так как имеет, грубо говоря, всего два свободно чередующихся выражения — тяжелый и большого веса — и потому для него в русском языке не нужна стандартная лексическая функция. Указанный смысл нарушает и требование (1) — он мыслим лишь при названиях мучных изделий — и требование (2), поскольку имеет только два выражения (черный и ржаной), и, стало быть, не является стандартной ЛФ.

Лексические функции обозначаются аббревиатурами, образованными от латинских слов. Так, функция, отвечающая смыслу 'очень', обозначается Magn — от лат. magnus 'большой'. Применяется следующая запись:

| Функция | Аргумент | Значение |
|---------|---------------------|--|
| Magn | (<i>брюнетка</i>) | = <i>жгучая</i> |
| Magn | (<i>рана</i>) | = <i>тяжелая</i> |
| Magn | (<i>знать</i>) | = <i>назубок; как свои пять пальцев... И Т. Д.</i> |

Среди стандартных ЛФ выделен подкласс простых ЛФ, которым противопоставляются сложные ЛФ; эти последние строятся из простых путем операции композиции функций.

Описание лексических функций:

Syn – синоним

Anti – антоним

Der – синтаксический дериват(учиться — учебный(другая часть речи))

Gener – обобщение(жидкое вещество - жидкость)

Sing – отдельный элемент множества(мак - зерно)

Mult – множество элемента (волк – стая)

Magn – высокая степень интенсивности(обыск — тщательный)

Ver – соответствующий назначению, истинный(догадка - верная)

Loc – место, локализация (завод – на заводе)

Oper – операция совершать(вывод – делать)

Func – функционирование(ветер — дуть)

Attr — параметр(офицер — погоны)

Plus, Minus — соответственно, 'более'/'менее' (маленький — крохотный)

Пример представления

| Входной параметр (слово) | Функция | Значение функции |
|-----------------------------|-------------|------------------------------|
| Элемент | Syn | часть |
| | Mult | множество |
| | Magn | множества |
| Множество | Syn | набор |
| | Sing | элемент |
| | Attr | мощность |
| | Loc | во множестве |
| Пересечение | Syn | скрещение |
| | Magn | множеств |
| | Func Loc | пересекать на пересечении |
| | Syn | эквивалентность |
| Равенство | Der | равные |
| | Magn | множеств |
| | | |
| Объединение | Syn | соединение |
| | Anti | разбиение |
| | Magn | множеств |
| | Der | объединять |
| | Func | объединять |
| Разбиение | Syn | разделение |
| | Anti | объединение |
| | Der | разбивать |
| | Func | разбивать |
| | Magn | множества |
| Принадлежность | Syn | иметь |
| | Gener | отношение |

| | | |
|--|-----|--------------|
| | | |
| | Der | принадлежать |

Перечень простых стандартных лексических функций с пояснениями см. в книге «Опыт теории лингвистических моделей «Смысл-Текст» Мельчука, с. 82-100.

Примеры статей ТКС, содержащие описание лексических коррелятов, см. там же, с. 113-130.