

Задание 5. RusVectores

Для определения отношений между словами в паре использовался тезаурус RuWordNet, чтобы исключить субъективность оценки этих отношений.

Слово	Ближайшее-1	Отношение	Ближайшее-2	Отношение	Ближайшее-3	Отношение	Ближайшее-4	Отношение	Ближайшее-5	Отношение
Голова	Башка	<i>синоним</i>	Спина	<i>когипоним</i>	Затылок	<i>комероним</i>	Шея	<i>когипоним</i>	Нога	—
Мяч	Мячик	<i>синоним</i>	Шайба	<i>когипоним</i>	Вратарь	—	Ракетка	<i>когипоним</i>	Пас	—
Герой	Персонаж	<i>гипероним</i>	Героиня	<i>синоним</i>	Сюжет	—	Роман	—	Повествователь	—
Шляпа	Шапка	<i>когипоним</i>	Картуз	—	Шляпка	<i>однокор.</i>	Фуражка	—	Треуголка	—
Диск	Пластинка	—	Компакт-диск	<i>синоним</i>	Альбом	—	Трек	—	Кассета	—

Достаточно часто первыми “ассоциациями” к слову являются просто формы этого слова, которые RuWordNet чаще всего относит к синсету этого слова, поэтому будем считать такие слова синонимами. Также среди них могут встретиться и когипонимы.

Когипонимы в целом встречаются чаще всего: среди первых пяти ближайших почти всегда найдется как минимум один (можно это наблюдать на нашей минимальной выборке). Среди первых трех распространены отношения синонимии, когипонимии и гиперонимии. Дополнительный анализ выявляет еще и комеронимию: например, голова и затылок согласно RuWordNet есть части некоторого целого — черепа.

Большое количество пар не имеет каких-то однозначно определенных отношений, если опираться на тезаурус, хотя с точки зрения “человеческой” оценки эти слова определенно имеют какую-то близость, их будем относить к related. Почти наверняка эти слова также являются когипонимами, но это происходит не напрямую, а транзитивно, то есть, чтобы найти их ближайшего общего предка, нужно подняться на несколько уровней вверх. Заметим, что в нашей таблице определенность начинает пропадать уже с четвертого ближайшего, в столбце 5 нет ни одного определенного прямого отношения, их близость RusVectores, скорее всего, “делает” из контекста, в котором два слова из пары встретились достаточно часто.

Можно провести дополнительный краткий анализ каждого слова в парах:

Голова

Голова — Башка: RuWordNet определяет их как синонимы, причем сразу в двух синсетах, составленных для двух разных лексических значений слова “голова”;

Голова — Спина: оба слова имеют общего предка “часть тела”;

Голова — Затылок: оба слова являются частями общего целого “череп”;

Голова — Шея: оба слова имеют общего предка “часть тела”;

Голова — Нога: слова не имеют общего предка, хотя кажется, что они относятся к “часть тела”.

Мяч

Мяч — Мячик: форма слова, RuWordNet также относит их к одному синсету;

Мяч — Шайба: имеют общего предка “спортивный инвентарь”

Мяч — Вратарь: являются близкими только благодаря частой встречаемости в контексте, например, текстов о спорте;

Мяч — Ракетка: имеют общего предка “спортивный инвентарь”;

Мяч — Пас: являются близкими по контексту.

Герой

Герой — Персонаж: связаны отношением “гипоним-гипероним”

Герой — Героиня: связаны общим синсетом, также являются формами слов друг друга;

Герой — Сюжет: являются близкими только по контексту;

Герой — Роман: не имеют отношений напрямую, близки по контексту;

Герой — Повествователь: не имеют отношений напрямую.

Шляпа

Шляпа — Шапка: имеют общего предка “головной убор”;

Шляпа — Шляпка: исключительно однокоренные слова;

Оставшиеся три слова, которые по сути определяют некоторый конкретный головной убор с единственным лексическим значением, близки только по контексту.

Диск

Диск — Компакт-диск: относятся к одному синсету;

Остальные ближайшие слова не имеют каких-либо прямых отношений со словом “диск”. Как можно заметить, большая часть из них дает нам некоторый “музыкальный” контекст, то есть RusVectores под диском подразумевает в первую очередь CD-диск, нечто более узкое, чем просто оптический носитель информации.