**Краткий анализ**

**1. norm\_pos\_ipm1\_10**

*Настройки:* нормализация L2, ipm > 1, 10 min/max, деление по частям речи, фильтрация по частям речи между словарём и моделью (последняя настройка в каждом файле, поэтому в следующих пунктах буду её опускать).

Глагол: например, в измерении 1 слова разделяются по категории ***негативное отношение (доминировать, дискредитировать, перераспределить, упорядочить) – положительное отношение (проститься, прощать, возлюбить, молить, помолиться…)***. 2-е измерение не интерпретируется однозначно. 3-е измерение: скорее всего, отношение ***шум (кипеть, бурлить, утопать, наболеть) – покой (советовать?, вознестись?)***. … 10-е: ***/тут сложно подобрать разумный гипероним/ – холод (зябнуть, сковать, покрывать, одеревенеть)***. 11-е. ***умственная работа (отгадать, перечислять, вспоминать, забываться) – работа руками (обтянуться, стилизовать, обглодать, обшить) = головастый – рукастый***… 151-е: ***свобода (отпустить, разгуляться, исчезнуть, отпускать) – неволя, труд (занимать, строить, метить, автоматизировать)***. 152-е: сложно интерпретировать… 241-е: ***вносить покой (обнадёживать, ободрять) – вносить смуту (схватить, заострить).*** Похоже на измерение 3, очень мало глаголов именно этой оппозиции.

Наречие: 1-е измерение: ***уверенно (прочно, законодательно, жестко, идеально, безукоризненно) – неуверенно (жаль, холодно)***. 2-е измерение: сложно дать интерпретацию. 3-е измерение: аналогично… 10-е измерение: ***тихо (молчком, втихую, секретно) – громко (бешено, отчаянно, нещадно, нестерпимо)***. 11-е измерение: сложно дать интерпретацию… НАРЕЧИЯ ИЗ-ЗА СВОЕГО НЕБОЛЬШОГО КОЛИЧЕСТВА ПОСЛЕ ФИЛЬТРАЦИИ РЕДКО ДАЮТ ХОРОШУЮ ИНТЕРПРЕТАЦИЮ!

Прилагательное: 1-е: ***дальний (господствующий, иерархический) – близкий (закадычный, приоткрытый). Мало прилагательных для интерпретации данной оппозиции***. 2-е: ***/сложно подобрать гипероним/ – довольный (сытый, довольный)***. 3-е: ***мирный (райский, обетованный) – вражеский (взрывчатый, сатирический, ружейный)***… 10-е: ***реальный (проверочный, обеспечительный) – нереальный (манный, несбывшийся, невозвратный, нематериальный)***. 11-е: сложно дать однозначную интерпретацию… 151-е: ***природный (августовский, ноябрьский, сентябрьский, июльский, декабрьский) – человеческий (хромосомный, чувашский, разговорный, русскоязычный)***. 152-е: ***из низших слоёв (низовой, инстинктивный?, дымный?, втянутый?) – из высших слоёв (дорогостоящий, гарвардский, экзотический)***… 241-е: сложно дать интерпретацию.

Существительное: 1-е: ***неблизкие отношения (народовластие, расслоение, господство) – близкие отношения (чад, чадо, кров, доча, возлюбленные)***. 2-е: ***правительство (конституционность, госкомитет) – народ (братуха, ткач)***. Мало существительных. 3-е: ***отсутствие проблемы (рай, лепта) – беды (инсулин, простатит, судимость, штамм, морфий)***… 10-е: сложно дать интерпретацию. 11-е: ***наука (фотосинтез, рецептор, биофизика, вирусология) – быт (лак, седан, замша, пена, уродина?)***… 20-е: ***наличие крепких связей (монокристалл, нуклетоид, дермантин, цинк) – слабые связи (слабина, прорубь, отпор?)***. 21-е: сложно дать интерпретацию… 29-е: ***болезнь (кдимакс, изжога, проседь, ревматизм, оскомина) – здоровье (здравие, молебен, святая (как сущ.))***. 30-е: ***сладкая жизнь (сахарница, сгущёнка, блюдечко?) – трудная жизнь (экстремизм, разжигание, комедия, новомученик…)***

**Предварительный вывод**: для интерпретации лучше брать ГЛАГОЛЫ и СУЩЕСТВИТЕЛЬНЫЕ, если у нас будет необходимость делить. При этом обращаем внимание, что оппозиционные темы для той или другой части речи различаются, но нечто общее выделить можно. Смотрите, например НЕГАТИВ – ПОЗИТИВ для измерения ГЛАГ-1 по настроению схоже с измерением НЕБЛИЗОСТЬ – БЛИЗОСТЬ СУЩ-1: люди с близкими отношениями друг к другу хорошо относятся, и наоборот. ГЛАГ-11: УМ – ТРУД, СУЩ-11: НАУКА – БЫТ. В науке люди работают, в основном, головой.

**2. ipm1\_plus\_pos\_10**

*Настройки:* нет нормализации, ipm > 1, 10 min/max, деление по частям речи…

Глагол: 1-е: та же самая оппозиция, но с другим набором лемм: ***упорочить, сгустить, дискредитировать – проститься, прощаться, чмокнуть, прощать, опомлиться***. 2-е: можно выделить ту же оппозицию, но она более неоднородная. 3-е***: быть горячим (кипеть, бурлить, закипать) – быть холодным (проветривать)***, мало слов… 10-е: ***/сложно сформулировать гипероним/ – быть холодным (зябнуть, сковать, овеять…)***. При дальнейшем рассмотрении было сложно выделить темы, а если темы и были, то они пересекались с уже существующими.

Наречие: 1-е: ***уверенно (прочно, законодательно, жестко) – неуверенно (голодно?, холодно)***. Тема перекликается с темой в нормализованном списке. 2-е: сложно дать интерпретацию. 3-е: ***неминуемость (вот-вот, неизбежно, непременно) – негативность (это вообще не оппозиция, но хоть что-то выделить смог: осуждающе, критически, неодобрительно)***… См. 9-е: слова БЕДА и ГРЕХ не могут быть наречиями, но в модели они были наречиями, ***оппозиция проблема (невнимательно, натощак…) – отсутствие проблем (смирно, расслабленно, отчётливо…)***. 10-е: ***чувственно (лично, понимающе, испытующе, удовлетворительно) – бесчувственно (пустынно, пусто, недвижно)***, 11-е: сложно дать интерпретацию, но, скорее всего, похоже на измерение 9. Наречия опять не дали явного описания.

Прилагательное: 1-е: ***дальний (господствующий, иерархический) – близкий (закадычный, приоткрытый, блаженный).*** Та же оппозиция, что и при нормализации. 2-е: ***проблемный (сейсмический, ультразвуковой) – довольный (те же слова)***. В ненормализованном варианте удалось подобрать название темы, хоть и приблизительное. 3-е: 3-е: ***мирный (райский, обетованный) – вражеский (взрывчатый, сатирический, ружейный)***… Та же тема. С прилагательным параметр нормализации практически не меняет набор слов.

Существительное: 1-е: ***неблизкие отношения (народовластие, расслоение, господство) – близкие отношения (чад, чадо, кров, доча, возлюбленные)***. Та же тема. 2-е: та же тема. 3-е: та же тема… 10-е: ***жизнь в городе (буллит, штатский (как сущ.), пленум, архив, резолюция) – жизнь загородом (конница, свежесть, кров, трапезная)***. 11-е: ***наука (фотосинтез, рецептор, биофизика, вирусология) – быт (лак, седан, замша, пена, уродина?)***… Та же тема.

**Предварительный вывод:** нормализация позволяет выделить большее количество оппозиций с большим количеством примеров для той или иной темы. Важно отметить, что во многих случаях набор слов остаётся почти неизменным.

**3. only\_ipm\_20**

*Настройки-1:* w2v\_sorted\_dimensions\_taiga\_20\_ipm10. 20 min/max, ipm > 10, без нормализации.

1-е: ***неблизкие отношения (мировой, иерархия, уставной, равенство, группировка) – близкие отношения (сестра, поцеловать, родные, блаженный, обнять, простить)***. Увеличение количества слов даёт больше тематических примеров, увеличение ipm уменьшает кол-во «шума». 2-е: ***объекты и действия неживой природы (имущественность, мебель, бинокль, перепись, звание, скорбь) – отношение к живой природе (сыграть, сытый, сказаться, хлопнуть)***. Оппозиция, конечно, очень «натянутая». 3-е: сложно дать определение оппозиции… 10-е: ***жизнь в городе (генеральный, пресс-служба, прокурор, законопроект) – жизнь загородом (цвести, влага, сияние, тело, кровь)***. 11-е: сложно дать оппозицию, хочется подобрать нечто связанное с питанием (см. min1-20). 12-е: ***нарушение прав (правоотношение, санкция, виновный, урегулирование) – соблюдение прав (трезвый, блеснуть, здравый)***.

*Настройки-2:* w2v\_sorted\_dimensions\_taiga\_20\_ipm7. 20 min/max, ipm > 7, без нормализации.

1-е: та же тематика. 2-е: та же тематика. 3-е: ***спокойствие (родной, правденый, райский, рай, литься, водопад) – суматоха (хлопнуть, боеприпас, патрон, антисоветский)***… 10-е: та же тема. 11-е: ***технологии и направления науки (метро, вирусология, приключения) – части технологий (двигатель, петля, топливный)***. 12-е: та же тема… Как вариант, предлагаю комбинировать ipm 7 и 10, т.к. тематический набор пока не меняется.

*Настройки-3:* w2v\_sorted\_dimensions\_taiga\_20\_ipm3. 20 min/max, ipm > 3, без нормализации.

Темы те же, тут нет смысла иллюстрировать особенности.

*Настройки-4:* w2v\_sorted\_dimensions\_taiga\_20\_ipm1. 20 min/max, ipm > 1, без нормализации.

1-е: та же тема, 2-е: вполне возможно, немного сместился аспект тематики: ближе, конечно же, к набору слов темы ***проблемы и покой***, но нет чёткой дифференциации по min/max. Например, для мин. слов с темой «покой» попадают негативно окрашенные слова (гудеть, пришить, жутко), есть слова и с позитивной окраской (по-хорошему, чмокнуть…). 3-е: см. старые темы. 4-е: ***труд физический (парикмахер, гарнитур, парфюмерия, пекарня, маникюр, портниха) – труд умственный (геохимический, классификация…)…*** 10-е: тема городской и загородной жизни не выделяется. 11-е: та же тема, что и с ipm > 3. 12-е: старая тема про ***нарушение прав***.

**3. norm\_ipm\_20**

*Настройки-1:* w2v\_sorted\_dimensions\_taiga\_20\_ipm10\_norm. 20 min/max, ipm > 10, нормализация L2.

1-е: ***неблизкие отношения (властный, иерархия, жестко, меньшинство, монополия, группировка) – близкие отношения (подруга, мать, внук, внучка, плач, муж, сестра)***. Тема та же, но БОЛЬШЕ тематических слов для оппозиции. 2-е: ***неживая природа (аппаратура, прибор, приемник, мебель) – живая природа (заяц, вена, слон, братец)***. Снова больше слов, эта тема теперь точно прослеживается. 3-е: ***спокойствие? (рай, море, первенство) – буря? (маршал, ездить, боеприпас, арсенал, патрон)***. Хотя в max слова попадают слова, связанные с беспокойством (ад, клуб и пр.)… 10-е: тут немного подумал, теперь мне кажется, что выделяется оппозиция ***работа (инспектор, хитрый, прокурор, архив, следователь) – отдых (красота, благодать, вечный, цвести…)***. Это кажется более разумным. 11-е: ***технологии и направления науки? (метро) – части технологий (подвеска, смесь, мотор)***… 17-е: ***мир (беседа, диалог, философ, приветствовать, философия) и вражда (враг, потерять, трогать, тыл)***. Темы интерпретируются лучше.

*Настройки-2:* w2v\_sorted\_dimensions\_taiga\_20\_ipm7\_norm. 20 min/max, ipm > 7, нормализация L2.

Те же темы, немного другие наборы слов.

*Настройки-3:* w2v\_sorted\_dimensions\_taiga\_20\_ipm3\_norm. 20 min/max, ipm > 3, нормализация L2.

Те же темы можно выделить.

*Настройки-4:* w2v\_sorted\_dimensions\_taiga\_20\_ipm3\_norm. 20 min/max, ipm > 3, нормализация L2.

В этих оппозициях намного меньше слов, которые мы выделяли при условии повышения ipm. Так, например, во втором измерении нет слов ***«слон», «заяц»*** и пр. Так что ipm > 1 нам не подходит, скорее всего.

***Общие выводы после просмотра всех остальных файлов (настройки)***.

1. Однозначно стоит смотреть на файлы с **нормализацией**.

2. При **ipm больше 7** отсеивается большая часть «шумных» слов.

3. Лучше брать **20 слов**, т.к. первая десятка может оказаться нерепрезентативной (1 тематическое слово или 0 слов).

4. **Деление на части речи иногда имеет смысл**, т.к. семантика групп может быть общей, но разные части речи выделяют разные аспекты (см. примеры на стр. 1–2). Упор делаем в таком случае на **ГЛАГОЛАХ и СУЩЕСТВИТЕЛЬНЫХ**.