

Содержание

Часть 2. Дистрибутивно-статистический анализ в микроинтервале (Статистическое открытие регулярной морфологии)	1
2.1. Позиции, n-граммы и парадигмы в статистическом открытии морфологии...	2
2.2. От протоаффиксов к парадигмам.....	7
2.2.1. Формирование протоаффиксов и протопарадигм	7
2.2.2. Формирование глагольных парадигм	9
2.2.3. Формирование адъективных парадигм	11
2.2.4. Формирование субстантивных парадигм	13
2.2.5. Уточнение парадигм	18
2.2.6. Префиксы	20
2.2.7. Парадигмы и размер исследуемого корпуса	23
2.3. Английский язык.....	25
2.4. Французский язык.....	32
2.5. Немецкий язык.....	36
2.6. Шведский язык.....	40
2.7. Итальянский язык.....	42
2.8. Латинский язык.....	44
2.9. Поиски парадигм – краткие итоги.....	50
Таблица 3.4. 100 самых частых существительных.....	54
Таблица 3.5. 100 самых частых прилагательных.....	55
Таблица 3.6. 100 самых частых глаголов.....	56

Часть 2

Дистрибутивно-статистический анализ в микроинтервале (Статистическое открытие регулярной морфологии)

В первой части книги (особенно п. 1.5.1, с. 54-55) были намечены пути формального открытия морфологии в пределах графического слова. Можно предложить множество подобных путей анализа, опирающихся на статистику комбинаторики букв в буквенных цепочках и на вероятность появления фрагментов таких цепочек в определенных окружениях.

В качестве первого примера рассмотрим подход, учитывающий статистику отдельных букв и буквосочетаний в их связи с началом и концом графического слова¹. Соответствующим корпусом послужил комплект "Независимой газеты" за 1996-2000 гг. (35 млн словоупотреблений). Из общего частотного словаря выделена зона частых слов с вероятностью ниже 0,001 и выше 0,000001 (всего более 62 тысяч разных графических слов).

2.1. Позиции, n-граммы и парадигмы в статистическом открытии морфологии

В общем случае исследователь в рамках ДСА опирается на буквы (с их частотами), на какие-то заданные позиции в тексте (например, рядом с пробелом) и буквенные цепочки (n-граммы). Все три рода данных связаны друг с другом. Иллюстрацией станет таблица 2.1.

Табл. 2.1.

Относительная частота букв (в промилле)

	Зона частых слов (0,000001 - 0,001)			Зона самых частых слов (>0,001)
	Начало слова	Внутри слова	Конец слова	
а	21	85	91	71
б	31	15	2	22
в	77	39	31	75
г	34	17	8	18
д	48	33	14	25
е	8	92	109	92
ж	7	10	1	20
з	28	15	5	15
и	34	81	98	98
й	...	4	85	-
к	59	31	12	40
л	19	46	18	26
м	49	23	63	22
н	64	77	16	76
о	67	117	91	109
п	148	13	1	20
р	54	65	12	16
с	122	58	11	45
т	26	75	50	68
у	24	22	31	13
ф	13	2	3	-
х	8	3	45	4
ц	6	8	1	-
ч	18	13	2	24
ш	5	7	...	-
щ	...	5	...	-
ъ	-	1	...	-
ы	-	17	43	19
ь	-	11	55	2
э	15	2	...	12

¹ См. Шайкевич, 2008.

ю	2	4	21	-
я	5	9	80	13

Анализ таблицы 2.1 приводит к двум результатам.

1) Относительные частоты букв могут сильно варьировать в разных зонах частотного словаря. В зоне самых частых слов частоты букв *б, в, ж, и, к, ч, э* существенно выше, чем в зоне частых слов. Объяснение приходит сразу же – отдельные слова с максимальной частотой сильно влияют на общую частоту соответствующих букв. Таковы предлог *в*, союз *и*, формы глагола *быть*, местоимений *этот, какой, который, что* и т. п. Учет этого вывода заставит задуматься о способах расчета безусловной вероятности той или иной буквы.

2) Почти все буквы (за исключением буквы *у*) резко меняют свою частоту в разных позициях в слове: в начале слова видим громадное повышение частоты букв *п, с, в, к, б, э, ф*; понижена частота *о, и, т, а*; минимальна частота *е* и *й*, а буквы *ъ, ы, ь* отсутствуют. В конце слова сосредоточены *й, я, м, ь, ы, х* и *ю*. Понятие позиции оказывалось ключевым в алгоритме Андреева. Условная вероятность какой-то буквы в данной позиции, деленная на безусловную вероятность, и была стартом при выявлении морфологических типов. Соответствующий коэффициент для конца слова был бы следующим: *й=5,7; х=4,5; я=4,2; ю=3,5; ь=3,4; ы=2,4; м=1,9*.

На современных компьютерах можно получать статистику буквосочетаний заданной длины — *n*-грамм. Подобные данные использовались в ДСА еще до обращения к задаче открытия морфологии. Речь идет о собственно графематическом анализе – получении дистрибутивных классов букв (ср. 1.4.4), возможному открытию графем-диграфов. Последняя задача едва ли важна для русского языка, но она может стать ключевой для таких языков, как польский или английский. В английском языке ДСА, возможно, обнаружил бы группу диграфов, не ограниченных позицией в слове (*ch, rh, sh, th*), *wh*, сконцентрированный в начале слов, *gh*, упрятанный внутрь слова, и собственно *h*.

Таблицы буквосочетаний были бы здесь слишком громоздки. Ограничимся демонстрацией неслучайных значений *S* у буквосочетаний *с о* (*p=0,106*), *п* (*p=0,031*) и *р* (*p=0,056*).

	S		S		S
оо	-1278	пп	-286	опр	+546
оп	-267	пр	+2908	орп	-149
ор	+356	ро	+828	пор	+340
по	+1294	рп	-504	про	+3487
		рр	-659	роп	+105
				рпо	-160

Ясно видны предпочитаемые буквосочетания и буквосочетания, встречающиеся ниже нормы.

Статистический анализ *n*-грамм может быть соединен с позицией. Взяв простейший случай диграмм, мы можем проверить степень концентрации данного буквосочетания в той или иной позиции. Так, у буквосочетания *ье* на конец слова приходится 99,6% всех случаев. Вот перечень других диграмм, у которых в конечной позиции появляется более половины всех случаев (против ожидаемых 13 в зоне частых слов): *-ии* 99,5%, *-шь* 99%, *-ый* 99%, *-ых* 96%, *-ях* 96%, *-ию* 93%, *-ой* 91%, *-ся* 90%, *-ют* 90%, *-сь* 89%, *-их* 87%, *-ия* 84%, *-ть* 83%, *-ая* 81%, *-ие* 77%, *-ий* 77%, *-ее* 74%, *-ым* 73%, *-ах* 69%, *-ей* 69%, *-ем* 67%, *-ую* 67%, *-ты* 64%, *-ое* 55%, *-им* 53%, *-то* 52%. Не столь высока концентрация диграмм в начале слова: *вз-* 95%, *пр-* 80%, *по-* 78%, *вс-* 68%, *за-* 57%, *вы-* 55%, *до-* 50%. Подавляющее большинство этих диграмм войдут в состав парадигм, но есть здесь и информационный шум: конечное *-ты* и начальное *пр-*.

Что касается буквосочетаний внутри слова, то их интерпретация должна быть отложена до позднейших этапов открытия морфологии. Некоторые из них станут частью морфем или основ, например *-бяз-* (*S*=+1370) в составе *обяз-*. Другие буквосочетания уже на этой стадии кажутся хорошими кандидатами в морфемы или комбинации морфем. Таковы *-вш-* (*S*=+300), *-йш-* (*S*=+460), *-ущ-* (*S*=+690), *-ющ-* (*S*=+2100), *-ств-* (*S*=+4900). Наконец, в некоторых случаях внутрисловное буквосочетание может оказаться сигналом морфемных чередований: *-чн-* (*S*=+200), *-жн-* (*S*=+380), *-ьн-* (*S*=+1200).

Основное направление дальнейшей работы – анализ n-грамм как потенциальных основ и их расширений как потенциальных аффиксов. Попробуем ввести ограничения на длину основ и аффиксов. Пусть минимальная длина основы будет равной четырем буквам, а максимальная длина аффикса не превышает половины длины основы. Вводя такой порог, мы на этом этапе отказываемся от анализа многих коротких основ (вроде *дом-ами, мо-ими, взя-лись*) в надежде получить регулярную парадигматику, а уже затем переходить к обработке коротких основ. Удобный материал для проверки левых основ содержится в прямом алфавитном словаре. Соответствующий обратный словарь дает нам перечень правых основ. Дополнения до целого графического слова у левых основ станут потенциальными суффиксами, дополнения правых основ будут префиксами. Так, у левой основы *весн-* будут найдены следующие продолжения (суффиксы): *-а, -е, -ой, -у, -ы*, в своей совокупности образующие **протопарадигму**. В соответствии с п. 1.5.1 будем называть **широтой** протопарадигмы число аффиксов, входящих в ее состав (в данном случае – 5). Число основ, обладающих данным набором аффиксов, будем называть **мощностью** протопарадигмы. В нашем случае мощность парадигмы равна 30 (*анкар-, атмосфер-, блокад-, борьб-, генпрокуратур-, годовщин-, госдум-, дружб-, европ-, екатерин-, ирин-* и т.д.). В том случае, когда основа совпадает с целым графическим словом, но при этом обнаруживаются и какие-то аффиксы, отсутствие продолжения объявляется нулевым аффиксом (0). Например, парадигма *-0, -го, -е, -й, -м, -му, -сти, -сть* обладает мощностью 20 (*актуально-, вольно-, действительно-, исключительно- ...*). У парадигмы *-0, -а, -ом, -у* обнаружено 46 основ (*арафат-, белград-, борт-, вадим-, гайдар-, гейдар-, генсек-, голод-, гражданин-, дьявол- ...*). Чем шире парадигма, тем больше ее морфологическая определенность, тем больше у нее шансов совпасть с традиционным представлением о парадигме. Так, парадигма *-0, -й* при достаточной мощности (30 основ) включает очень разнородный материал (*безгранично-, богаче-, буре-, дружно-, иваново-, иронически- ...*). Заметим, что полученные парадигмы покрывают не только традиционно понимаемое словоизменение, но и словообразование. Особенно ясно это видно у префиксальных парадигм.

Префиксальные парадигмы отличаются широтой, но их мощность может быть весьма значительной. Вот важнейшие из них (с указанием мощности):

<i>0-, не-</i>	333	(<i>-велика, -малое, -которыми, -уверенности ...</i>);
<i>0-, по-</i>	222	(<i>-крова, -важнее, -иске, -гибли, -чести ...</i>);
<i>0-, с-</i>	70	(<i>-высока, -начала, -можете, -умели ...</i>);
<i>0-, со-</i>	65	(<i>-стояла, -братьев, -хранение, -жгли ...</i>);
<i>0-, на-</i>	49	(<i>-всегда, -завтра, -звании, -печатали ...</i>);
<i>0-, за-</i>	48	(<i>-явка, -хотела, -будьте, -купили ...</i>);
<i>0-, у-</i>	48	(<i>-гроза, -знала, -видите ...</i>);
<i>0-, гос-</i>	43	(<i>-сектора, -чиновников, -комитете, -премий ...</i>);
<i>0-, про-</i>	31	(<i>-демонстрировала, -исходила, -шествия ...</i>);
<i>0-, рас-</i>	27	(<i>-продажа, -скажите, -положении ...</i>).

Конечно, в полученных парадигмах присутствует информационный шум в виде пар правых основ, ошибочно объединенных в рамках одной парадигмы, ср. *рема/зарема, града/награда, даре/ударе, лице/улице, грамма/программа, алии/реалии*, но доля таких ошибок невелика. Среди маломощных парадигм встречаются и полностью фиктивные, ср. парадигму *0-, д-* с парами основ *остатка/достатка, вора/двора, оклад/доклад, ушами/душами, ревности/древности*. Однобуквенность префикса увеличивает вероятность ошибки. Уникален случай парадигмы *0-, с-*, где доля ошибок превышает 50%, ср. *танка/станка, тройка/стройка, тулу/стулу, цене/сцене*. Начальные консонантные сочетания со звуком *с-* столь обычны в индоевропейских языках, что приписывать им морфологический статус было бы рискованно.

Число суффиксальных парадигм очень велико. Десять из них имеют мощность более 199, у 52 парадигм мощность превышает 25 основ. Рекордсменом выступает парадигма *-0, -а* с 872 основами (*азот-, бахтин-, вальс-, глазунов-, дебют-, есенин-, живот-, зайцев-, имам-*). Особого внимания заслуживают две мощные парадигмы с единственным маркированным суффиксом: *-0, -ся* с 578 основами (*берут-, вернем-, встречать-, имеющими- ...*) и *-0, -сь* с 162 основами (*вернула-, готовили-, заняло- ...*). Оба суффикса встречаются в составе более широких парадигм, например, в парадигме *-0, -а, -ась, -и, -ись, -о, -ось, -ся* (*встречал-, высказал-, использовал-, изменил- ...*). Спаренность данных

суффиксов, их графическое сходство позволяют считать их вариантами одной морфемы, присоединяемой к некоторым парадигмам.

Изредка полученные парадигмы точно соответствуют каким-то традиционным типам словоизменения. Такова парадигма -0, -а, -ам, -ами, -ах, -е, -ов, -ом, -у, -ы (аспект-, вариант-, взгляд-, вопрос-, документ-, журнал-, институт- ...) и почти точная -0, -а, -ам, -ах, -е, -ой, -у, -ы (войн-, газет-, групп-, жертв- ...).

Чаще полученные парадигмы представляют собой фрагменты традиционных парадигм, например, субстантивных: -и, -ов (баск-, венк-, жулик-, значк- ...); -и, -ь (акварел-, важност-, ерес-, жидкост- ...); -а, -о (бешенств-, божеств-, гнезд-, древ- ...); -ей, -и, -ю, -я (абхазы-, бюрократы-, грецы-, ирони- ...); -е, -ем, -и, -ю, -я (введени-, внимани-, давлени-, закрыти- ...). Прообразами глагольных парадигм выступают -0, -т (бегу, веря-, делаю-, ищу- ...); -л, -ла, -ли, -ть (висе-, воева-, вылете-, заговори- ...); -ет, -ли, -ть, -ют (влия-, добыва-, запреща-, исчеза- ...). Наконец, примерами фрагментов адъективных парадигм служат -ий, -ого (атлантическ-, белинск-, вольск-, жванецк- ...) и -ой, -ые, -ых (агитационн-, войсков-, заграничн- ...).

В традиционном представлении адъективные парадигмы обладают наибольшей широтой, понятна поэтому малая вероятность найти в тексте полный набор соответствующих окончаний. Так, основа *иностранн*- в зоне частых слов имеет следующие продолжения: -ая, -ого, -ое, -ой, -ому, -ую, -ые, -ый, -ым, -ыми, -ых, для завершения парадигмы полного прилагательного не хватает лишь -ом. Основа *возглавляем*- имеет те же продолжения за исключением -ыми. Переход к зоне слов средней частоты привел бы к тому, что эти две основы оказались бы в пределах одной парадигмы.

Но где границы искомых парадигм? Ведь рядом с парадигмой основ типа *иностранн*- существуют парадигмы с дополнительными суффиксами -а, -о, -ы (возможн-, адекватн- и пр.). Расширение зоны поиска до слов средней частоты приводит к тому, что все больше основ проявляют сочетаемость с этими суффиксами (*избирательна, временна, глобальна; денежно, информационно*). Более того, в парадигму включаются новые суффиксы: *бессмысленнее, адекватнее, естественней, актуальней*. Лишь суффикса -ен не хватает до полной традиционной парадигмы прилагательного (за вычетом суперлатива).

Существует еще одна мощная адъективная парадигма с исходом на -ск-, где на месте буквы *ы* в суффиксах мы находим букву *и* (*германск-, исламск-* и т. д.). Вместо суффикса -о обнаруживаем -и у основ *идеологическ-* и *демократическ-*. Переход к зоне средней частоты увеличивает круг подобных наречий (*европейски, граждански*), но мы не найдем здесь ни кратких прилагательных, ни компаратива. Попытка расширить парадигму оказалась бы тщетной, в русском таких форм нет.

Понижение порога частоты приводит к постоянному расширению парадигмы, наращиванию ее мощности, к слиянию разных парадигм. Если в зоне частых слов полная традиционная парадигма существительных среднего рода с исходом на -ств- ограничивалась 9 основами (*ведомств-, государств-, достоинств-, качеств-, министерств-, обществ-, правительств-, убийств-, чувств-*), то теперь их число выросло до 32 (*банкротств-, законодательств-, меньшинств-, существ-* и т. п.). Вместе с тем многие парадигмы не расширяются до теоретической полноты не из-за вероятностных ограничений, но вследствие каких-то семантических факторов. Действительно, именно семантика ответственна за то, что не выявлены формы множественного числа у таких слов, как *бегство, большинство, варварство, взяточничество, воровство, господство, дворянство, детство, духовенство, единство, животноводство, изыщество, лидерство* и т. п.

Абстрактность существительного мешает его употреблению во множественном числе. Так, в зоне частых слов были обнаружены мощнейшая (хотя и узкая) парадигма с суффиксами -и и -ь (около 200 основ, большей частью с исходом на -ост) и близкая парадигма с суффиксами -и, -ью и -я. Соответствующая полная парадигма засвидетельствована только у основ: *возможност-, договоренност-, област-, обязанност-, особенност-, способност-, ценност-*. Понижение частотного порога приводит к переходу большинства основ из первой парадигмы во вторую, т. е. к достраиванию списка форм единственного числа. Но лишь 20% из них пополняются формами множественного числа (*акварел-, банальност-* и т. п.).

В зоне частых слов выделилась парадигма с суффиксами -ей, -и, -й, -ю, -я, -ям, -ями, -ях, она включает 20 основ (*инстанци-, монополи-, ситуаци-, фотографи-* и т. д.). Понижение порога частоты резко увеличивает мощность

парадигмы до 90 основ (делегаци-, интерпретаци-, конституци-, презентаци-, цивилизаци- и т. п.). Но некоторые основы с суффиксами -ей, -и, -ю, -я отказываются строить формы множественного числа. В первую очередь это названия стран абхаз-, араби-, англи-, албани-, венгри-, греци-, ингушети-, индонези-, испани- и т. д. К ним же примыкают и некоторые имена нарицательные: авиация, агитация, документация, интеллигенция, инфляция, ирония и т. п.

Построению полных парадигм могут мешать исходно принятые ограничения на соотношение длины основы и длины аффикса. Например, в только что рассмотренной парадигме присоединению суффикса -ями препятствует краткость основ лекци-, лини-, парти-, пенси-, преми-, сери-, сесси-, студи-, теори-, эмоци-.

В связи с этим возникает еще один путь уточнения парадигм: принимая как промежуточный итог парадигмы, полученные в зоне частых слов, попробуем снять указанные ограничения на длину основы. В зоне частых слов с суффиксом -ами соединятся основы автор-, актер-, выбор-, глаз-, завод-, игрок-, идеал- и многие другие. У правой основы -делил к первоначальным префиксам у-, вы- присоединится раз-, а с понижением порога частоты также 0-, на-, по-, от-. Соответственно у основы -любил к префиксам 0-, по- прибавятся раз- и воз-. Уже в зоне частых слов левая основа дава- присоединит -йте, -лись и -лось к прежним суффиксам -й, -л, -ла, -ли, -ло, -ть, -я.

Оба пути уточнения парадигм (т. е. переход к зоне слов средней частоты и снятие ограничений на длину основ и аффиксов) сопровождаются появлением новых парадигм, связанных преимущественно со словообразованием. У вполне интегрированных в адъективную парадигму основ вооруженн-, глобальн-, естественн-, избирательн- появляются суффиксы -ости и -ость, ведущие к субстантивным парадигмам. Субстантивный суффикс -ик обнаруживается у адъективных основ духовн- и временн-. К субстантивным основам присоединяются не только аффиксы адъективных парадигм: институт- + -ские, -ских, -ской; инвалид- + -ной, -ном, -ную, -ные, -ных, но и субстантивные аффиксы: вопрос-, институт- + -ик; газет-, голов-, групп-, икон- + -ка; австри- + -йцы. В зоне частых слов основа вспомина- соединена с глагольными суффиксами -ем, -ет, -ется, -ешь, -л, -ла, -ли, -ть, -ю, -ют, -ются, понижение частотного порога добавляет глагольный суффикс -ете, адъективный -вших и субстантивный -ние. Возникают особые субстантивные парадигмы (оппозиции мужского и женского) с суффиксами -к, -ка, -ца, -цы (основы великомучени-, волшебни-, воспитанни-, выпускни-, грешни-, заложни-, защитни- и т. п.) и с суффиксами -ь, -я, -ница, -ницы (основы воспитател-, деятел-, жител-, зritel- и т. п.). Постоянное "размножение" парадигм при понижении порога частоты создает все новые и новые трудности, в общих чертах путь их преодоления представляется следующим.

В зоне частых слов формируются потенциальные автономные аффиксы из тех узких парадигм, где положительный аффикс противопоставлен нулевому. Для суффиксов такими внешними аффиксами кажутся -сь и -ся; среди префиксов наиболее яркий кандидат не-. Затем все графические слова частотного словаря переписываются заново уже без указанных внешних аффиксов ("частиц"). Тем самым несколько сокращается широта и разнообразие парадигм, увеличивается их мощность.

Снова определяются парадигмы в обеих частотных зонах, но при этом особое внимание обращается на вид основ при аффиксах конкретной парадигмы. Если эти основы кончаются (или начинаются) на определенную букву (или комбинацию букв), то выдвигается гипотеза о компонентах этих основ, иначе говоря, о существовании основ разного порядка. Так, взяв только что приведенный пример с парадигмой -ь, -я, -ница, -ницы, находим все основы, соединяющиеся хотя бы с тремя из этих суффиксов. Таких основ оказалось 36, пять из них принадлежат к зоне частых слов: жител-, исполнитель-, писател-, представител-, учител-, совпадая во всех четырех суффиксах. Сразу бросается в глаза асимметрия двух пар суффиксов. Частота суффикса -ница — 661, -ницы — 419, другие суффиксы из той же субстантивной парадигмы: 0 — 181, -ам — 22, -ами — 9, -е — 61, -ей — 149, -у — 90 (отсутствует суффикс -ах из полной парадигмы типа больница). Частота суффикса -ь — 27364, -я — 20308, другие суффиксы из той же субстантивной парадигмы: -е — 322, -ей — 28353, -ем — 7937, -и — 21835, -ю — 2943, -ям — 4466, -ями — 6720, -ях — 328 (т. е. полная субстантивная парадигма типа автомобиль). Последняя парадигма почти в 100 раз превышает первую по совокупной частоте.

От основ с исходом на -ль- создаются основы двух адъективных парадигм: 1) зрительск-, издательск-, исполнительск-, любительск-, писательск-, предпринимательск-, председательск-, представительск-, преподавательск-,

приятельск-, свидетельск-, учительск- и читательск- с общей частотой 2790; 2) воспитательн-, деятельн-, зрительн-, исполнительн-, настоятельн-, основательн-, победительн-, представительн-, целительн- с общей частотой 8870.

От основ с исходом на -льн- образуются субстантивные основы исполнительность-, основательность-, представительность- и исключительно частая основа деятельность- (частота = 18891)².

Непосредственно с основами на -тель- (основами первого порядка) соотносятся субстантивные основы с исходом на -ств- (основы второго порядка): жительство-, издательств-, исполнительств-, любительств-, писательств-, покровительств-, предпринимательств-, представительств-, свидетельств-, сожительств-, учительств- с совокупной частотой 9188. Особое положение занимает основа правительств-, чья частота составляет 52873. С ним соотносятся две адъективные основы третьего порядка: правительственн- (5823) и правительствующий (25).

От некоторых основ с исходом на -ств- (покровительств-, председательств-, представительств- и свидетельств-) образованы по две глагольные основы третьего порядка с исходом на -ова- (с общей частотой 661) и с исходом на -у- (1391), от последних созданы адъективные основы четвертого порядка с исходом на -ющ- (517).

Что касается самих основ на -тел-, то многие из них сами восходят к глагольным основам исследова-, воспита-, жи-, изда-, исполни-, исследова-, любь-, надзира-, носи-, облада-, основа-, обита-, писа-, победи-, почита-, прави-, предпринима-, преподава-, представи-, руководи-, служи-, смотри-, созда-, учи-, храни-, цели- и чита-. В остальных случаях явная первооснова в рамках ДСА не просматривается (благодетел-, властител-, деятел-, зритель-, настоятел-, обывател-, покровител-, председател-, приятел-, свидетел-, сожител-).

2.2. От протоаффиксов к парадигмам

2.2.1. Формирование протоаффиксов и протопарадигм

Существует множество путей статистического открытия морфологической структуры внутри графического слова. Тот путь, который был описан в предыдущем параграфе, учитывал частоту буквенных цепочек в корпусе текстов. Вернемся к нашему основному корпусу прозы и проверим еще один метод выявления парадигм. На этот раз зависимость от частот слов в тексте будет учитываться только в виде некоторого порога частоты, вводимого для ограничения списка графических слов, на котором будет работать программа формирования парадигм. Установление такого порога существенно сокращает вычислительную работу. Если в нашем корпусе встретилось 15 графических слов, начинающихся с буквенной цепочки емельян (емельян, емельяна, емельяна-то, емельянихи, емельяниху, емельянов, емельянова, емельяновке, емельяновой, емельянову, емельяновы, емельяном, емельяну, емельянушка, емельяныч), то лишь три слова преодолели установленный нами порог частоты (8): емельян – 61, емельяна – 22, емельяну – 9. Теперь емельян участвует в работе алгоритма на тех же правах, что и иван, частота которого в тексте в сто раз больше.

Снова устанавливаются пороги для длины основы (минимально 4 буквы) и для аффикса (максимально 4 буквы). Как только какая-то буквенная цепочка допускает хотя бы два продолжения в списке графических слов, начинается формирование разных основ. Начальная цепочка емел имеет только одно продолжение – ь, но цепочка емель допускает два продолжения – к и я, первое продолжение емельк, в свою очередь, допускает три продолжения: а, и, у (все они – конечные буквы графического слова). Цепочка емельк становится основой, а три продолжения объявляются протоаффиксами, их набор считается протопарадигмой. Ту же протопарадигму /-а, -и, -у/ находим при основах женк-, жженк-, заметк-, зденк-, земляник-, золовк-, илюшечк-, кокетк-, конфетк-, копеечк-, косичк-, косушк- и т. п. (всего – 71 основа). Несколько шире протопарадигма /-а, -ами, -и, -ой, -у/ с основами загадк-, занавеск-, игрушк-, нагайк- (всего – 10 основ); еще шире протопарадигма /-а, -ами, -е, -и, -ой, -у/ с основами закуск-, копеечк-, крышк-

² Заметим, что речь идет о формальных соотношениях основ: деятель – деятельный – деятельность. Реальное направление словопроизводства не может быть определено без привлечения семантических данных.

(всего – 14 основ). Протопарадигма /-а, -ам, -ами, -ах, -е, -и, -ой, -у/ с основами *выходк-*, *записк-*, *повозк-* почти совпадает с традиционно понимаемой парадигмой, отсутствует лишь форма род. п. мн. ч. (-ок, -ек). В рамках семантической лингвистики отождествление всех этих вариантов достигается моментально; в ДСА предстоит еще долгий путь.

Буквенная цепочка *емелья* допускает лишь одно продолжение -н, на этом, с одной стороны, завершается графическое слово и постулируется нулевой протосуффикс, а с другой – протосуффиксами становятся два других продолжения а и у. Возникшая протопарадигма /-0, -а, -у/ относится к числу самых мощных – 177 основ.

Такой алгоритм породил более 10 тысяч протопарадигм, около 35 тысяч основ, 734 протосуффикса.

Результаты машинной обработки выдаются в следующих форматах:

1. Парадигмы с основами, например,

берег

1 0 а0 ам0 ами0 ах0 е0 и0 ись0 ите0 ла0 ли0 ов0 ом0 у0 ут0

правил

1 0 а0 ам0 ами0 ах0 е0 и0 о0 ом0 у0 ьная0 ьнее0 ьно0 ьное0 ьной0 ьном0 ьною0 ьную0 ьны0 ьные0 ьный0 ьным0 ьных0

сапог

1 0 а0 ам0 ами0 ах0 е0 и0 ов0 ова0 ом0 у0

жених извозчик мальчик мужик начальник овраг помещик праздник старик успех чиновник

11 0 а0 ам0 ами0 ах0 е0 и0 ов0 ом0 у0

бумаг собак

2 0 а0 ам0 ами0 ах0 е0 и0 ой0 ою0 у0

2. Частотный словарь протосуффиксов, например,

18201	0	37	ав0	14	аете0
26	а-то0	25	аем0	13	аешь0
6686	а0	226	ает0	58	ай0

3. Пары протосуффиксов, входящих в одни и те же протопарадигмы:

	Число общих парадигм	Число основ в этих парадигмах		Число общих парадигм	Число основ в этих парадигмах
0 а	1542	5152	0 ской	64	65
0 ам	273	338	0 сь	74	1519
0 ах	347	426	0 ся	162	2318
0 и	450	3723	0 те	158	496
0 ми	132	1456	0 у	798	1945
0 ов	494	859	0 ы	846	1235
0 ом	700	1176			

4. Значения S для пар протосуффиксов, входящих в одни и те же протопарадигмы:

S		S		S		S	
324	0 а	21	0 ли	37	0 ный	80	0 те
44	0 ам	280	0 м	25	0 ным	31	0 тся
62	0 ами	367	0 ми	27	0 ных	147	0 у
174	0 ась	66	0 му	77	0 о	304	0 х
49	0 ах	27	0 н	87	0 ов	32	0 ши
128	0 го	23	0 на	85	0 ом	31	0 ший
197	0 е	29	0 ная	120	0 ось	30	0 шка
49	0 есь	21	0 ного	60	0 сти	85	0 ы
318	0 и	21	0 ное	62	0 сть	24	0 ым
20	0 ии	33	0 ной	41	0 стью	71	0 ь
137	0 ись	20	0 ну	409	0 сь	44	0 ься
23	0 ия	21	0 ную	526	0 ся	73	0 ю
143	0 й	31	0 ные	82	0 т	35	0 я

Пары протосуффиксов могут быть ранжированы по значениям их S:
1695 ми х 388 е м 243 в м 198 ув ул

1069	ась	ся	386	й	ми	241	л	ть	197	ь	ься
1024	м	х	367	0	ми	226	есь	ься	196	вне	вны
842	ась	ись	364	го	м	226	й	ю	194	а	ась
792	м	ми	349	есь	ся	220	го	сть	194	ул	уть
766	ись	ся	328	го	е	220	сти	стью	193	ли	ть
758	ась	ось	328	сти	сть	218	го	сти	190	ула	уть
730	й	х	324	0	а	215	ул	ула	188	ые	ых
626	й	м	320	е	ми	212	сть	стью	186	вне	вну
589	ись	ось	318	0	и	212	ся	ься	186	ла	ть
585	го	му	304	0	х	210	и	ись	185	ула	ули
583	е	х	290	а	и	209	и	ся	185	ые	ый
537	го	й	280	0	м	205	л	ла	185	ь	ью
526	0	ся	276	й	му	202	те	шь	184	и	ось
517	ось	ся	270	м	му	201	ув	ула	182	ув	уть
512	е	й	258	а	у	200	ла	ли			
409	0	сь	254	ась	и	198	л	ли			

2.2.2. Формирование глагольных парадигм

Легко видеть, что уже на самом верху этого ранжированного списка пары протоаффиксов группируются в кластеры – обобщенные парадигмы. Четыре протоаффикса *-ась*, *-ись*, *-ось*, *-ся* самым естественным образом объединяются, ведь все они связаны друг с другом (среднее $S=755$). Весь комплект этих четырех протоаффиксов представлен у 62 основ, одну из этих основ используем как ярлык для краткого обозначения соответствующей парадигмы, пусть это будет БОРОЛ-ся. Еще 159 основ сочетаются с тремя из данных четырех аффиксов. Наконец, около 50 основ соединились со всеми 4 данными аффиксами, но также с какими-то другими аффиксами. Все полученные основы кончаются на *-л* и, вероятно, вместе с другими основами образуют сложную глагольную систему. Заметим, что многие из этих основ всегда сопровождаются показателем *ся/сь* и образуют особый класс медиальных глаголов, ср. *виднел*, *ворвал*, *годил*, *касал*, *мерещил*, *нравил*, *остал*, *осунул*, *откликнул*, *очутил*, *понадобил*, *появил*, *разгорал*, *смеял*, *столпил*, *толпил*, *улыбнул* и т. п.

Двигаясь вниз по ранжированному списку, мы обнаружим новые значимые связи: вышеприведенные аффиксы коррелируют с *0*, *-а* и *-и*. Создается парадигма ВОДИЛ, где 70 основ присоединяют все 7 аффиксов (и никаких других), а несколько сот основ сочетаются с 5–6 аффиксами из этого набора, либо присоединяют какой-то аффикс вне набора. Характерно сочетание рассматриваемых семи аффиксов с *-о* (полторы сотни основ типа *беспокоил-*, *вырвал-*, *потерял-*, *считал-* и т. п.). Последний протоаффикс не вошел в семичленный набор из-за многозначности русской морфемы *-о* (семантика среднего рода у существительных и форм прошедшего времени глагола, но также семантика наречия). Матрица значений S вновь присоединенных аффиксов и четырех аффиксов с *-ся/сь* (обозначена -СЯ) выглядит следующим образом:

	-СЯ	-0	-а	-и	-о
-СЯ	755	239	158	214	98
-0	239		324	318	77
-а	158	324		290	155
-и	214	318	290		99
-о	98	77	155	99	

Своеобразно поведение конечного *-ся*. Помимо того, что он входит в две предыдущие парадигмы (БОРОЛ-ся и ВОДИЛ), он появляется как конечный компонент в парадигме */-аяся, -еися, -ейся, -иися, -ийся, -ихся, -уюся/* НАЧАВШ/КАСАЮЩ (более 30 основ, среднее значение $S = 80$). Те же основы (без *-ся*) входят в парадигмы адъективного типа, но одновременно связаны с глагольными основами. Это заставляет признать *-ся* частицей или таким особым производным аффиксом, который не противостоит другим маркированным аффиксам³. Подобное же поведение характерно и для префикса *не-*.

³ С общелингвистической точки зрения *-сь* выступает как алломорф *-ся* после гласного во всех глагольных формах, кроме причастий.

Две первые парадигмы содержали основы с конечным -л; это заставляет подозревать, что -л – какой-то суффикс, а наши парадигмы – лишь конечные компоненты в многослойной парадигме. Действительно, движение вниз по ранжированному списку S скоро обнаружит -л как протоаффикс в новой более широкой парадигме /-в, -ет, -ется, -л, -ла, -лась, -ли, -лись, -ло, -лось, -лся, -ть, -ться, -ют, -ются/. Вот матрица попарных S внутри этой парадигмы⁴:

	-в	-ет	-ется	-л	-ла	-ли	-ло	-ЛСЯ	-ть	-ться	-ют	-ются
-в		6	-	113	115	88	45	41	93	31	5	-
-ет	6		77	113	97	101	51	37	113	34	148	54
-ется	-	77		53	43	45	23	81	56	85	71	126
-л	113	113	53		205	198	97	79	241	72	97	37
-ла	115	97	43	205		200	116	71	186	63	86	32
-ли	88	101	45	198	200		109	69	193	62	99	35
-ло	45	51	23	97	116	109		46	92	36	46	18
-ЛСЯ	41	37	81	79	71	69	46	125	80	128	35	74
-ть	93	113	56	241	186	193	92	80		82	98	41
-ться	31	34	85	72	63	62	36	128	82		32	76
-ют	5	148	71	97	86	99	46	35	98	32		57
-ются	-	54	126	37	32	35	18	74	41	76	57	

Бросается в глаза одно «аномальное место» в этой матрице – слабая связь -в с -ет, -ется, -ют и -ются, легко объясняемая с позиций семантической лингвистики: -в – это показатель прошедшего времени в деепричастиях, а -ет – флексия настоящего времени.

Нелегко подобрать ярлык для этой парадигмы. Нет ни одной основы, которая присоединяла бы весь набор аффиксов. Пусть представителем парадигмы будет НАЗЫВА-ть: основа *называ-* присоединила 14 аффиксов (кроме -в) и еще 12 аффиксов, чья связь с рассматриваемой парадигмой обнаруживалась бы по мере снижения S⁵): -ем (12), -емое (12), -емые (15), -емый (15), -емых (10), -ете (40), -ешь (29), -й (18), -йте (36), -ю (29), -юсь (30), -я (49). Если рассмотреть все остальные обнаруженные протоаффиксы, то у 13 из них среднее S с аффиксами парадигмы НАЗЫВА-ть превысит 10: -вшая (17), -вши (19), -вшие (22), -вший (29), -вших (17), -емся (14), -йся (20), -тесь (14), -ющие (11), -ющий (16), -ющим (11), -ющих (13), -ясь (48).

Так постепенно определяются контуры того, что в семантической лингвистике называется глагольной парадигмой: выявились личные окончания, показатели инфинитива и деепричастий, наметились формы трех причастий. Исключительная широта складывающейся парадигмы (40 обнаруженных аффиксов) объясняет низкую мощность агрегируемых парадигм, причиной неполноты набора аффиксов у конкретной основы может быть ее малая частота. Самая мощная протопарадигма с 32 основами (*взира-*, *запуска-* и пр.) имеет только 4 общих аффикса (-ет, -л, -ть, -я); 14 общих аффиксов находим лишь у одной пары основ: *отвеча-* и *поступа-*. Заметим, что в рассматриваемой парадигме нет аффикса -0. Это указывает на связанный характер производящей основы (большинство основ заканчивается буквой а).

Спускаясь по ранжированному списку, мы, действительно, обнаружим парадигму, где основы уже не оканчиваются на а, напротив, у многих аффиксов это а обнаруживается. Данная парадигма ПРОЧИТ-ать включает /-ав, -аем, -ает, -аете, -ай, -айте, -ал, -ала, -али, -ался, -ан, -ана, -ано, -аны, -ать, -аю, -ают, -ывай, -ывал, -ываю, -ывая/. Матрица внутренних связей свидетельствует о некоторой неоднородности парадигмы⁶:

	АЕТ	АЛ	ать	АН	ЫВ
АЕТ	65	55	66	4	2
АЛ	55	141	120	36	18
ать	66	120		42	20

⁴ Под символом ЛСЯ объединены -лась, -лись, -лось, -лся.

⁵ В скобках указывается среднее S в связях аффикса с 12 аффиксами рассматриваемой парадигмы.

⁶ Символом ЫВ обозначены четыре аффикса с этим буквосочетанием; соответственно используется символ АЛ; символ АН объединяет -ан, -ана, -ано, -аны и -ав; -ать показан отдельно; все остальные аффиксы объединены под символом АЕТ.

АН	4	36	42	73	28
ЫВ	2	18	20	28	78

С семантической точки зрения здесь объединился довольно разнородный материал как глаголы совершенного вида (сказать, рассказать), так и производные от них глаголы несовершенного вида, оформленные по-разному (рассказывать, забывать, встречать), отыменные глаголы несовершенного вида (работать).

С двумя последними парадигмами сходна парадигма СТРЕЛ-ять /-аем, -яет, -яете, -яй, -ял, -яла, -яли, -яло, -ялся, -ять, -яю, -яют, -яя/ со средним значением $S=74$ у внутренних связей.

Общее конечное -н основы характеризуют парадигму ТОЛКН-уть /-ет, -ув, -ул, -ула, -ули, -уло, -ут, -уть/ со средним значением $S=104$. Небольшой набор аффиксов обеспечивает достаточно мощную парадигму из шести основ: блесн-, вздрогн-, вспыхн-, выглян-, дрогн-, толкн-. Ясно, что должна появиться и парадигма типа НАМЕК-нуть /-нет, -нув, -нул, -нула, -нут, -нуть/ со средним значением $S=53$.

Некоторым аналогом парадигмы НАЗЫВА-ть можно считать парадигму ПУСТИ-ть /-0, -в, -л, -ла, -ли, -ло, -м, -т, -те, -шь/, в которой подавляющее большинство основ кончается на букву и. Две самые мощные протопарадигмы содержат по 5 основ (выпусти-, допусти-, отложи-, посади-, пропусти- с 9 аффиксами и кричи-, молчи-, посиди-, прогони-, услыши- с 5 аффиксами).

В какой-то степени вариантом той же парадигмы можно считать парадигму ГОВОР-ить /-ив, -ил, -ила, -или, -ило, -ит, -ить, -ишь, -ят/ со средним значением $S=85$.

Подобно тому, как в парадигме ПРОЧИТ-ать выявился компонент -ан, -ана, -ано, -аны, рядом с парадигмой ГОВОР-ить появляются парадигмы: ОПРЕДЕЛ-ЕН /-ен, -ена, -ено, -ены/ со средним значением $S=85$ и ВЫСТАВ-ЛЕН /-лен, -лена, -лено, -лены/ со средним значением $S=117$. Традиционная лингвистика моментально указала бы на закономерность: вариант с -л появляется после губных согласных. ДСА может лишь перечислить эти согласные (-б, -в, -м, -п), но не может обобщить результат, сказав, что это губные.

На парадигмы НАЗЫВА-ть и ПРОЧИТА-ть похожа парадигма СКРИП-еть /-ел, -ела, -ели/ со средним значением $S=121$.

Особое положение занимает парадигма ЧУВСТВ-овал /-овал, -уем, -ует, -уете, -уй, -уют, -уя/ со средним значением $S=64$.

Последняя парадигма, которую следует здесь упомянуть, включает полный вариант /-0, -ка, -те/ с 34 основами: возьми, вспомни, выпей, выпьем, гляди, давай, изволь, налей, погляди, погоди, подай, подойди, подумай, поедem, пожалуй, позволь, пойдем, поживи, покажи, полно, попробуй, послушай, посмотри, стой, пошли, принеси, прочти, расскажи, скажи, слушай, смотри, спроси, ступай, сходи и редуцированный вариант /-0, -ка/ с 14 основами: велите, глянь, дайте, подите, подумайте, пойдемте, пойду, покажь, попробуйте, стойте, пусть, расскажите, садись, садитесь.

Сходство рассмотренных до сих пор парадигм, проявляющееся в буквенном совпадении многих аффиксов, позволяет сформировать большой класс парадигм, который традиционная лингвистика назвала бы глагольным. Достижение и уточнение парадигм будет производиться после снятия некоторых ранее введенных ограничений (порог частоты, пороги длины основы и аффикса), но черновое определение глагольного класса удалось получить в рамках графического слова. Вряд ли этот класс серьезно перестроится на минимальном интервале ДСА.

2.2.3. Формирование адъективных парадигм

Вернемся к началу ранжированного списка пар протоаффиксов. В пятерке первых пар две пары сыграли свою роль в формировании глагольных парадигм, оставшиеся три протоаффикса -м, -ми и -х будут встречаться во многих парадигмах, но никакой эксклюзивной⁷ протопарадигмы они не создали. Восьмое и девятое место в ранжированном списке занимает -й, и тут же появляется первая адъективная

⁷ Будем называть парадигму эксклюзивной, если она включает все те и только те аффиксы, которые явным образом перечислены в наборе данной парадигмы.

трехчленная протопарадигма /-й0, -м0, -х0/ с основами артиллерийски-, верноподданны-, высочайши-, десяти-, еврейски- ... всего 35 основ.

Опустившись до S=500, обнаружим еще три протоаффикса -го, -е, -му, которым суждено войти во многие адъективные парадигмы. Появляется первая эксклюзивная протопарадигма с пятью протоаффиксами /-го, -е, -й, -м, -му/: английско-, барско-, ближайше-, будуще-, бывше-, военно- ... всего 68 основ. Затем следует четырехчленная эксклюзивная протопарадигма /-е, -м, -ми, -х/: беспрестанны-, боковы-, висящи-, гнилы-, голубы- ... всего 51 основа. Почти эксклюзивны пятичленные протопарадигмы /-го, -е, -й, -м, -му, -ю/: большо-, верхне-, весенне- ... всего 45 основ; /-0, -го, -е, -й, -м, -му/: беспокойно-, блаженно-, божественно-, временно-, высочайше-, далеко-, душевно-, жарко- ... всего 28 основ.

Дойдя до S=250, увидим, что -го и -ми тесно связаны с протоаффиксом -0. Складывается первая агрегированная парадигма с восемью аффиксами /-0, -го, -е, -й, -м, -ми, -х/, пусть ее ярлыком будет ГОЛОДНО. В составе этой парадигмы 396 основ с шестью аффиксами из данного набора, 666 основ с пятью аффиксами, 1092 с четырьмя аффиксами. Было бы преждевременным считать ее адъективной, лучше назвать ее именной. Дело в том, что одна мощная парадигма /-0, -м, -ми, -х/ носит явно именной характер; из ее 270 основ лишь одна (епанчины) относится к адъективам, все остальные – субстантивы (города, господа, греха, гусара, движения, двора и т. п.).

Когда значения S достигают величины 120, множество протопарадигм легко агрегируется в широкую и мощную адъективную парадигму ТЯЖЕЛ-ый⁸ /-ая (122), -ого (126), -ое (133), -ой (124), -ом (77), -ому (101), -ою (85), -ую (125), -ые (137), -ый (131), -ым (129), -ыми (111), -ых (121)/. Тринадцать аффиксов создают такую широкую парадигму, что шансов обнаружить эксклюзивную парадигму практически нет. Основа ТЯЖЕЛ-, выбранная в качестве ярлыка, кроме этих 13 аффиксов присоединяют еще пять аффиксов -а, -ее, -ей, -о, -ы. Такой набор почти соответствовал бы полной парадигме традиционной лингвистики. Кроме тяжел- еще 89 основ сочетаются с этими 13 аффиксами: благородн-, бледн-, богат-, болезненн-, быстр-, весел-, внезапн-, вольн-, главн-, громадн-, грустн-, грязн-, действительн-, деревянн-, длинн-, добродушн-, духовн-, единственн-, естественн-, законн-, замечательн-, значительн-, знаком-, известн-, интересн-, испуганн-, истинн-, казенн-, каменн-, котор-, красив-, красн-, кругл-, ласков-, любезн-, любим-, мертв-, мрачн-, народн-, наружн-, невинн-, недостойн-, неизвестн-, некотор-, необыкновенн-, неожиданн-, непонятн-, неприятн-, несчастн-, ничтожн-, нравственн-, обыкновенн-, обычн-, огромн-, особенн-, открыт-, печальн-, подобн-, полезн-, постоянн-, почтенн-, прекрасн-, приятн-, разбит-, раскрыт-, решительн-, светл-, свободн-, семейн-, серебрян-, серьезн-, сильн-, скверн-, скромн-, собственн-, спокойн-, старинн-, странн-, страстн-, страшн-, суров-, счастлив-, таинственн-, тверд-, толст-, холодн-, церковн-, честн-, четверт-.

Любопытны внешние связи этой главной адъективной парадигмы с некоторыми внешними аффиксами: -о 90, -ее 61, -ость 55. Намечаются контуры более широкой парадигмы. Напомним, что в основе дистрибутивно-статистического выделения парадигм лежит идея регулярности (в данном случае – в пределах графического слова). Регулярные морфологические связи выходят за пределы традиционных представлений о частях речи. Если формы компаратива традиционно включаются в парадигму прилагательного, формы на -о попадают туда же лишь в функции предикатива, приглагольное или приадъективное употребление объявляется особым словом – наречием. ДСАТ скорее включит все формы на -о в единую адъективную парадигму подобно тому, как деепричастия обычно включаются в единую глагольную парадигму.

Интересен вопрос о суперлативе. При заданных порогах на длину протоаффикса образовалась парадигма БЛАГОРОДНЕ-йший /-йшая, -йший, -йшим, -йших/, не показавшая связей с адъективными парадигмами. В состав этой парадигмы входят 23 основы, сочетающиеся, по крайней мере, с двумя аффиксами из четырех: благородне-, богате-, важне-, верне-, главне-, дальне-, добре-, искренне-, любезне-, огромне-, отличне-, перве-, полне-, почтенне-, превосходне-, прекрасне-, прелестне-, сильне-, совершенно-, старе-, счастливе-, ужасне-, чисте-. Шестнадцать из них сочетаются с аффиксами -ее и -о. При допущении

⁸ В скобках указывается среднее S в связях аффикса с остальными аффиксами рассматриваемой парадигмы.

пятибуквенных аффиксов (и с переходом -е в аффикс) вся эта группа усилит свою мощь до 40 (великодушн-, глуп-, изящн-, подробн- и т. п.) и целиком войдет в главную адъективную парадигму.

Следующая выделяющаяся парадигма ХОРОШ-ий /-ая (74), -его (90), -ее (57), -ей (61), -ем (32), -ему (79), -ею (24), -ие (82), -ий (89), -им (91), -ими (85), -их (109), -ую (79)/ напоминает предыдущую хвостами аффиксов и заменой *ы* на *и*, *о* на *е*. Все 13 аффиксов представлены только у основы горяч-. Абсолютное большинство основ этой парадигмы – это шипящие согласные -ж-, -ш-, -ч-, -щ- (например, будущ-, рабоч-, стоявш-, ближайш-, всеобщ-, похож-). Напомним, что эти буквы не появляются перед -ы-, -ю-, -я-, следовательно, их групповая близость проявляется уже в графематическом анализе в таблице бинарных буквосочетаний. Впрочем, существуют 22 основы с исходом на -н с двумя заменами аффиксов -ая на -ая, -ую на -ую: верхн-, весенн-, вечерн-, внешн-, внутренн-, домашн-, древн-, замужн-, здешн-, излишн-, искренн-, крайн-, нынешн-, передн-, последн-, посторонн-, прежн-, соседн-, средн-, тамошн-, тогдашн-, утрENN-.

Еще одна сходная парадигма ВЫСОК-ий /-ая (97), -ие (70), -ий (81), -им (80), -ими (72), -их (91), -ого (105), -ое (85), -ой (91), -ом (63), -ому (80), -ою (98), -ую (104)/. Основы как правило в конце имеют -к (изредка -г). Графематические ограничения у -к и -г – те же, что и в предыдущей парадигме. Многие основы этой парадигмы (например, близк-, крепк-, горьк-, сладк-, долг-, строг-) к перечисленному набору добавляют -а-, -и-, -о-. Многими сотнями измеряется число основ с исходом на ск.

Разные исходы основ не изменяют исходов перечисленных адъективных аффиксов. С позиций семантической лингвистики такое поведение свидетельствует о существовании морфем. Рассмотрим в качестве примера случай парадигмы ХОРОШ-ий. Начнем с основ, заканчивающихся на -ш-. Формально выделяемый аффикс -шая сочетается с основами *вы/пришед-, за/порос-, погиб-, привык-,* но в очень многих случаях (около 100) -шая следует за буквой *в-*; значит, -вшая есть соединение суффикса -вш- и флексии -ая. Разобранный выше аффикс -йший также позволяет выделить флексию -ий, но где левая граница суффикса? -й- или -ей-? Что делать с основами *ближайш-, величайш-, высочайш-, глубочайш-, жесточайш-, кратчайш-, крепчайш-, мельчайш-, низчайш-, строжайш-, тончайш-*? Здесь открываются сложные правила морфемных чередований *к/ч, г/ж, лк/льч* и т. п. Парадигма /-ная, -ного, -ное, -ной, -ную, -ные, -ный, -ным, -ных/ допускает разбиение на -н- и -ая (ср. *народ-, вздор-, комнат-, разум-, случай-, телес-*). Парадигма /-чная, -чно, -чное, -чной, -чную, -чные, -чный, -чным, -чных/ разбивается на -н-ая и предшествующую основу с исходом на -к/ч (ср. *восточный*), *г/ж* (ср. *бумажный*), *х/ш* (ср. *грешный*). Другое решение для случая -ьяная, где *ь* отойдет к основе (ср. *капитальная, правильная, похвальная*). Но что делать с /-нная, -нно, -нное, -нной, -нную, -нные, -нный, -нным, -нных/? Этим аффиксам всегда предшествует гласная буква, и значит, могут существовать -енная, -анная, -янная. Частично это отъемные прилагательные (мужественная, бессмысленная, болезненная, письменная, деревянная), частично – образования от глаголов парадигмы ГОВОР-ить (*благословенная, брошенная, возмущенная, вооруженная, воспитанная, выдуманная, вызванная, потерянная* и т. п.).

2.2.4. Формирование субстантивных парадигм

Максимальные внутренние связи (среднее $S=253$) показывает протопарадигма /-сти, -сть, -стью/, к которой явно примыкают /-стей, -стию, -стям, -стями, -стях/. Создается первая агрегированная парадигма с восемью аффиксами, к ней относится более двухсот основ, из которых лишь 8 имеют на конце *е* (*добле-, живуче-, пове-, поче-, преле-, свеже-, сове-, тяже-*), все остальные оканчиваются на *о*. Выберем в качестве ярлыка ОПАСНО-сть (с семью аффиксами); по 6 аффиксов у *благодарно-, должно-, обязанно-, способно-*, по 5 у *любезно-, местно-, неприятно-, подробно-, потребно-*. Почти все основы (98%) входят в адъективные парадигмы.

Еще одну агрегированную парадигму образует тройка аффиксов /-ции, -цию, -ция/ со средним значением $S=142$, к ней примыкает аффикс -тор ($S=91$). Ярлыком парадигмы назовем АДМИНИСТРА-ция, четыре аффикса присоединяются также к основам *опера-, редак-, экзеку-*. По три аффикса находим у основ *агита-, депута-, дирек-, инквизи-, корреспонден-, планта-, провин-* (!), *протек-* (!), *репети-*.

Третья явная агрегированная парадигма («отчества») АЛЕКСЕЕ-вич имеет восемь аффиксов /-вич, -вича, -вичу, -вна, -вне, -вной, -вну, -вны/ со средним значением S=142. Полностью восемь аффиксов видны еще у восьми основ: алекса́ндро-, андре́е-, василье́-, миха́йло-, николае́-, платоно́-, семено́-, степано́-. Общее число основ не превышает 50. Внешние связи (среднее S=52) с -ва и -ву ведут нас к парадигме («фамилии») НЕКРАС-ов /-ов, -ова, -ову, -овым/ со средним значением S=69. Все 4 аффикса видны у основ григор-, кирилл-, литвин-, меркул-, неведом-, неждан-, сверст-, степан-, стукол-, хвостик-, черемис-, щелкал-. Чаше встречаются основы с тремя первыми аффиксами: базар-, васильк-, волын-, давид-, денис-, долгору́к-, ермол-, замет-, зосим-, колос-, кутуз-, лопух-, маркел-, молот-, молчан-, мороз-, найден-, налет-, новосел-, облом-, пален-, полоз-, раскольник-, розан-, рязан-, салтык-, семен-, сидор-, симон-, скворц-, стебельк-, устин-, федор-, христ- (!), чума́к-.

Многие сотни и даже тысячи основ принадлежат парадигме ОБРАЗ /-0, -а, -ам, -ами, -ах, -е, -ов, -ом, -у, -ы/, которая объединяет множество протопарадигм. Все десять аффиксов представлены у 33 основ: взгля́д, вопро́с, генера́л, дива́н, докуме́нт, журна́л, зако́н, идеа́л, интере́с, карма́н, клас́с, миллио́н, наряд, огоро́д, отве́т, офице́р, пере́ход, порыв, предме́т, прие́м, приме́р, принци́п, промы́сл, пункт, разгово́р, расче́т, резу́льтат, рома́н, сове́т, стака́н, столб, тала́нт, францу́з. По 9 аффиксов видим у основ: бере́г, биле́т, выво́д, выстре́л, губерна́тор, гуса́р, забо́р, коридо́р, литерато́р, магази́н, наряд, нуме́р, о́бряд, о́стров, солда́т, студе́нт. Внутренние связи аффиксов показаны в следующей матрице.

	-0	-а	-ам	-ами	-ах	-е	-ов	-ом	-у	-ы
-0		324	44	62	49	197	87	85	147	85
-а	324		73	104	79	163	116	140	258	169
-ам	44	73		135	111	60	79	57	86	54
-ами	62	104	135		120	71	119	87	105	77
-ах	49	79	111	120		86	78	56	90	59
-е	197	163	60	71	86		53	66	174	82
-ов	87	116	79	119	78	53		112	83	85
-ом	85	140	57	87	56	66	112		100	66
-у	147	258	86	105	90	174	83	100		118
-ы	85	169	54	77	59	82	85	66	118	

Вариантом этой парадигмы можно считать набор аффиксов, где вместо -ы появляется -и: ЖЕНИХ /-0 (146), -а (172), -ам (68), -ами (98), -ах (87), -е (112), -и (139), -ов (88), -ом (83), -у (130)/. Эксклюзивной парадигмой здесь выступает набор основ: жени́х, извозчи́к, мальчи́к, мужи́к, нача́льник, овра́г, помещи́к, праздни́к, стари́к, успе́х, чи́новник. Как видим, многие перечисленные основы имеют исход на -ик. Естественно, что можно ожидать появление парадигмы ПРАЗДН-ик /-ик, -ика, -икам, -ике, -ики, -иков, -иком, -ику/ со средним значением S=81. По восемь аффиксов имеют основы нача́льн-, поко́йн-, свяще́нн-, чи́новн-; по 7 – воспита́нн-, преступн-, противн-, работн-, родственн-; по 6 – исправн-, любовн-, памя́тн-, страна́нн-; по 5 – истери́к- (явная ошибка!), помещи́к-, посла́нн-, послу́шн-, пусты́нн-, современн-, шко́льн-. Некоторое сходство с парадигмой ПРАЗДН-ик демонстрирует парадигма ПОСЛОВ-иц /-иц, -ица, -ице, -ицей, -ицу, -ицы/ со средним S=105. По шести аффиксов находим у основ воспита́нн-, гостин-, корми́л-, красав-, любовн-, работн-, родственн-; по пяти – больн-, висе́л-, лестн-, маслен-, помещ-, сестр-; по 4 – масте́р-, пуго́в-, страна́нн-, страна́нн- (всего более 60 основ).

Хвосты аффиксов этой парадигмы ведут нас к чрезвычайно мощной (более тысячи основ) парадигме ПРИЧИН-а /-а, -е, -ой, -ою, -у, -ы/. Эксклюзивную парадигму находим в случае основ: бездн-, белизн-, борьб-, быстрот-, вышин-, глубин-, доброт-, дружб-, ивановн-, княжн-, корон-, николаевн-, нищет-, павловн-, простот-, пустот-, темнот-, теплот-, тишин-, толщин-, чистот-, шири́н-. Полный набор аффиксов наблюдаем также у 48 основ: бесе́д-, боро́д-, вели́чин-, грамо́т-, доса́д-, екатери́н-, жало́б-, женщи́н-, жертв-, забо́т-, за́щит-, зве́зд-, каре́т-, катери́н-, комна́т-, красот-, литерату́р-, мину́т-, моги́л-, моли́тв-, москв-, мужчи́н-, награ́д-, наде́жд-, натур-, невест-, оград-, пере́мен-, персон-, полно́т-, полови́н-, польз-, посуд-, похва́л-, правд-, приро́д-, просьб-, работ-, сва́дьб-, свобо́д-, сестр-, сигар-, служб-, сторон-, страна́нн-, судьб-, угро́з-, фигу́р-.

«Мягкой» параллелью к парадигме ПРИЧИН-а можно считать парадигму КОНЮШ-ня /-не, -ней, -ни, -ню, -ня/: барыш-, герои-, господ-, дерев-, колоколь-, пусты-, святы-, сплет-.

Мощные парадигмы (ОБРАЗ, ЖЕНИХ, ПРИЧИН-а) – аналоги мужского и женского склонения традиционной русистики.

У аналогов слов среднего рода раньше всего обнаруживается парадигма ЗНАКОМ-ство /-а, -е, -о, -у/ со средним значением S=118. По четыре аффикса находим у основ: беспокой-, благород-, блажен-, государ-, дворян-, достоин-, наслед-, началь-, ничтоже-, обстоятель-, покровитель-, правитель-, производ-, простран-, самоубий-, совершен-, товарище-, устрой-, христиан-, человеке-. Довольно слабы основания для выделения парадигмы ДЕРЕВ /-о, -ом, -у/. Все три аффикса сочетаются с основами: болот-, войск-, гнезд-, зеркал-, колес-, кресл-, личик-, окошечк-, окошк-, письм-, правил-, ремесл-, стекл- (не считая основ на -ств-).

Среднее значение S=114 имеет парадигма ЗАКЛЮЧ-ение /-ение, -ении, -ений, -ению, -ения/. Все пять аффиксов наблюдаем у основ: заключ-, наблюда-, направл-, наслад-, объявл-, объясн-, определ-, оскорбл-, предлож-, предполож-, представл-, приглас-, приготовл-, произвед-, раздраж-, размышл-, разреш-, распоряж-, рассужд-, соображ-, стремл-. Такое же среднее значение имеет парадигма ВОСПИТ-ание /-ание, -анию, -ания/, полностью представленная образов-, оправд-, послуш-, предпис-, приказ-, содерж-, толков-. Обобщением этих двух парадигм становится парадигма /-ние, -нии, -ний, -нию, -ния/ и еще дальше парадигма ДЕЙСТВ-ие /-ие, -ием, -ию, -ия, -иям, -иями, -иях/ со средним значением S=83. Полный набор семи аффиксов видим также у 15 основ воспоминан-, выражен-, движен-, намерен-, наслажден-, описан-, показан-, предан-, предположен-, рассужден-, сомнен-, сочинен-, страдан-, убежден-, увлечен-. Нет ничего удивительного в конечном н основ, поэтому ниже перечислим лишь основы с другим исходом: занят-, извест-, несчаст-, понят-, услов- (по 6 аффиксов); предчувств-, прилич-, путешеств-, следств- (по 5 аффиксов); благополуч-, благород-, великодуш-, добродуш-, евангел-, милосерд-, насил-, невер-, открыт-, отсутств-, препятств-, приветств-, присутств-, проклят-, простодуш-, противореч-, развит-, разнообраз-, соглас-, сочувств-, усерд- (по 4 аффикса).

Уже при значении S=128 обнаружена пара -ье – -ья (бездел-, богомол-, варень-, вязан-, гулян-, жалован-, забыт-, зареч-, кушан-, мужич-, нездоров-, ненаст-, печен-, повер-, подпол-, помест-, предмест-, раздум-)⁹. Теперь выясняется, что эта пара ассоциируется с парадигмой ДЕЙСТВ-ие (среднее S=100), самые характерные основы воспоминан-, движен-, намерен-, наслажден-, рассужден-, сомнен-, страдан- получают теперь вариант с -ье. Возрастает и число доказанных основ: бесчест-, весел-, воскресен-, лечен-, прощан-, сиден-, соглас-.

Среднее значение S=128 показала одна узкая парадигма, соответствующая словам среднего рода, ЧУДОВ-ище /-ища, -ище/: обитал-, пепел-, пожар-, пристан-, прозв-.

Около S=100 в ранжированном списке появляется четырехчленная парадигма СОЧИНИ-тель /-тели, -тель, -телю-, -теля/, полностью представленная у основ заседа-, покупа-, посети-, представи-, следова-, смотри-, управи-.

Обобщением этой парадигмы можно считать ПРОФИЛ-ь /-ем, -ь, -я/ с эксклюзивной парадигмой: вензел-, мишел-, обладател-, поручител-, ридикюл-. Все три аффикса сочетаются также с основами богатыр-, гвозд-, карас-, крендель-, пастыр-, спектакл-, тюлен-, щегол-.

Опускаясь дальше по ранжированному списку, впервые сталкиваемся с чередованием букв внутри формирующихся парадигм. Примером может служить парадигма БУТЫЛК-а/ЖЕЛУД-ок /-ка, -ке, -ки, -ку, -ок/ со средним S=102. Все пять аффиксов встретились у 49 основ: арестант-, бесед-, бутыл-, волос-, вывод-, выдум-, выход-, голов-, город-, гувернант-, дворян-, десят-, желуд-, закус-, запис-, казач-, картин-, короб-, крестьян-, лошад-, малют-, мещан-, недостат-, останов-, палат-, перелес-, переул-, перчат-, повоз-, подбород-, поезд-, покуп-, попыт-, поступ-, прогул-, продел-, распис-, решет-, сирот-, скрип-, служан-, сосед-, справ-, старич-, стрел-, тарел-, тетрад-, уступ-, цыган-. По 4 аффикса у основ: англичан-, бабен-, беспоряд-, вывес-, задат-, замет-, занавес-, корот-

⁹ Но здесь же встретились основы игумен-, настас-, наталь-, попад-, устин-, девич-, охотнич-, собач-, человек-, которые не обнаружат общих аффиксов с парадигмой ДЕЙСТВ-ие.

(!), мужич-, налив-, напиток-, недоим-, остат-, отрыв-, перед-, перепел-, песен-, печен-, подар-, поедин-, поправ-, посел-, постав-, посыл-, потолок-, пригор-, припад-, промежут-, проступ-, рассуд-, склад-, товар-, торгов-, участ-, холод-, шарман-; по 3 аффикса – берез-, богомол-, булав-, ботин-, голуб-, достат-, дурач-, жаворон-, заработ-, избыт-, институт-, кредит-, образ-, околот-, отговор-, перекрест-, петров-, пирож-, потом-, почин-, приживал-, просел-, сверт-, червяч-.

Данная парадигма связана одновременно с парадигмами ЖЕНИХ и ПРИЧИН-а, что проявляется в ее внешних связях с аффиксами -кам, -ками, -ках, -ков, -кой, -ком (в среднем S=56).

Близок случай парадигмы НАСМЕШ-ка /-ек, -ка, -ке, -ки, -ку/ с основами бабуш-, бумаж-, верхуш-, девоч-, девуш-, дорож-, карточ-, лавоч-, мальчиш-, насмеш-, отлуч-, палоч-, подуш-, привыч-, рубаш-, собач-, старуш-, строч-, фураж-. По 4 аффикса у основ денеж-, игруш-, казач-, клетуш-, кусоч-, местеч-, платоч-; по 3 – у годоч-, листоч-, человек-. Аффикс -ек связан (S=33)¹⁰ с четырехчленной парадигмой ВАСЕН-ька /-ька, -ьке, -ьки, -ьку/ (среднее S=94), где по 4 аффикса имеют основы варен-, дяден-, полин-, наден-, постел-; по 3 – барын-, деревен-.¹¹

На парадигмы ЖЕЛУДОК и НАСМЕШ-ка похожа парадигма СКВОР-ец /-ец, -ца, -цы/ (среднее S=96): американ-, безум-, выход-, глупов-, голштин-, гостин-, жереб-, живопис-, знаком-, иностран-, испан-, итальян-, колод-, красав-, любим-, мерзав-, младен-, молод-, незнаком-, образ-, ополчен-, ординар-, очевид-, питом-, полковод-, послан-, преображен-, продав-, проходим-, румян-, сослужив-, счастлив-, ташкент-, торгов-, червон-, чечен-.

Подобно парадигме ВАСЕН-ька засвидетельствована парадигма /-ьца, -ьцев, -ьцы/ (среднее S=86) с тремя основами владел-, земледел-, постоял-.

С семантической точки зрения 25 основ, относящихся к парадигме /-ин, -ина, -ину/ (среднее S=103), должны распасться, по крайней мере, на три кластера: 1) велич-, развал-, старш-; 2) бабушкин- и многие фамилии: истом-, красотк-, ракит-, рогож-, рябин-, солом- и др.; 3) англичан-, господ-, граждан-, дворян-, крестьян-, мещан-, татар-. Формальное представление об этом дает матрица связей кластеров с некоторыми протоаффиксами.

	Кластер 1	Кластер 2	Кластер 3
-иной	66	5	-
-ины	170	65	-
-ами	-	-	68
-е	-	15	70

С формальной точки зрения особой парадигмой (хотя и с низким значением S) следует признать объединение двух парадигм с 12 аффиксами /-ская, -ские, -ский, -ским, -ских, -ское, -ской, -скую, -ства, -стве, -ство, -ству/. Среднее значение в первой парадигме S=120, во второй – S=118, связи двух парадигм имеют среднее S=20. Число соответствующих основ невелико¹², но все они крайне любопытны семантически – христиан-, человеке- (12), дворян- (11), господ-, граждан-, старче- (10), щеголь- (9), мастер- (8), варвар- (6), королев-, плутов- (5).

Подведем итог пункту 2.2. Список пар протоаффиксов, ранжированный по величине S, т. е. по степени статистической неслучайности, стал основой для кластеризации протопарадигм. На множестве десяти тысяч протопарадигм сформировались 43 агрегированные парадигмы. Ниже дается их список, где слева указан ярлык, а справа – аффиксы парадигмы:

Глагольные парадигмы

¹⁰ Примером соответствующих основ служат васил-, ступен-, козыр-, кошел-, флигел-.

¹¹ Преобладание н в исходе этих основ заставляет думать о суффиксе -ньк-, эта гипотеза подтвердится после снятия ограничения на длину основ и аффиксов (см. п. 3.2.5), ср. субстантивные основы агашеньк-а, алешеньк-а, анниньк-а, афоньк-а, бабеньк-а, батеньк-а, васеньк-а и др.; адъективные основы аленьк-ий, бедненьк-ий, беленьк-ий, веселеньк-ий и т. п.

¹² В скобках указано число обнаруженных аффиксов.

1. НАЗЫВА-ть	/в, вшая, вши, вшие, вший, вших, ем, емое, емся, емые, емый, емых, емся, ет, ете, ется, ешь, й, йся, йте, л, ла, лась, ли, лись, ло, лось, лся, тесь, ть, ться, ю, юсь, ют, ются ющие, ющий, ющим, ющих, я, ясь/
2. ВОДИЛ	/0, а, и, о/
3. ПРОЧИТ-ать	/ав, аем, ает, аете, ай, айте, ал, ала, али, ался, ан, ана, ано, аны, ать, аю, ают, ывай, ывал, ываю, ывая/
4. СТРЕЛ-ять	/аем, яет, яете, яй, ял, яла, яли, яло, ялся, ять, яю, яют, яя/
5. ПУСТИ-ть	/0, в, л, ла, ли, ло, м, т, те, шь/
6. ГОВОР-ить	/ив, ил, ила, или. ило, ит, ить, ишь, ят/
7. ОПРЕДЕЛ-ен	/ен, ена, ено, ены/
8. ВЫСТАВ-лен	/лен, лена, лено, лены/
9. ТОЛКН-уть	/ет, ув, ул, ула, ули, уло, ут, уть/
10. НАМЕК-нуть	/нет, нув, нул, нула, нут, нуть/
11. СКРИП-еть	/ел, ела, ели/
12. ЧУВСТВ-овал	/овал, уем, ует, уете, уй, уют, уя/
13. ГЛЯДИ	/-ка, 0, те/
14. БОРОЛ-ся	/ась, ись, ось, ся/
15. НАЧАВШИЙся/КАСАЮЩ-ийся	/аяся, еesia, ейся, иesia, ийся, ихся, уюся/

Адъективные парадигмы

16. ГОЛОДНО	/0, го, е, й, м, ми, х/
17. ТЯЖЕЛ-ый	/а, ая, ее, ей, о, ого, ое, ой, ом, ому, ою, ую, ы, ые, ый, ым, ыми, ых/
18. ХОРОШ-ий	/ая, его, ее, ей, ем, ему, ею, ие, ий, им, ими, их, ую /
19. БЛАГОРОДНЕ-йший	/йшая, йший, йшим, йших/
20. ВЫСОК-ий	/ая, ие, ий, им, ими, их, ого, ое, ой, ом, ому, ою, ую/

Субстантивные парадигмы

21. ОБРАЗ	/0, а, ам, ами, ах, е, ов, ом, у, ы/
22. ЖЕНИХ:	/0, а, ам, ами, ах, е, и, ов, ом, у/
23. ПРАЗДН-ик	/ик, ика, икам, ике, ики, иков, иком, ику/
24. СКВОР-ец	/ец, ца, цы/
25. ПРОФИЛ-ь	/ем, ь, я/
26. СОЧИНИ-тель	/тели, тель, телю, теля/
27. ПРИЧИН-а	/а, е, ой, ою, у, ы/
28. КОНЮШ-ня	/не, ней, ни, ню, ня/
29. ПОСЛОВ-иц	/иц, ица, ице, ицей, ицу, ицы/
30. НАСМЕШ-ка	/ек, ка, ке, ки, ку/
31. ВАСЕН-ька	/ька, ьке, ьки, ьку/
32. БУТЫЛК-а/ЖЕЛУД-ок	/ка, ке, ки, ку, ок/
33. АЛЕКСЕЕ-вич	/вич, вича, вичу, вна, вне, вной, вну, вны/
34. ОПАСНО-сть	/стей, сти, стию, сть, стью, стям, стями, стях/
35. ДЕРЕВ	/о, ом, у/
36. ЗНАКОМ-ство	/а, е, о, у/
37. ЧУДОВ-ище	/ища, ище/
38. ДЕЙСТВ-ие	/ие, ием, ию, ия, иям, иями, иях/
39. ЗАКЛЮЧ-ение	/ение, ении, ений, ению, ения/.
40. ВОСПИТ-ание	/ание, анию, ания/
41. АДМИНИСТРА-ция	/тор, ции, цию, ция /
42. НЕКРАС-ов	/ов, ова, ову, овым/
43. ВЕЛИЧ, ИСТОМ, МЕЩАН	/ин, ина, ину/

В некоторых парадигмах (24, 30, 32) обнаружены беглые гласные, замечены регулярные чередования конечных букв основы (к/ч, г/ж, х/ш) и начальных букв аффиксов (парадигмы 7 и 8).

Следует заметить, впрочем, что в русских существительных ограниченный набор флексий и вытекающая отсюда частая омонимия затрудняют разделение многих морфологических типов. Не удалось распознать типы *герой, прокофий, статуя, линия, горе*; исключительно редко краткие прилагательные включались в адъективные парадигмы; из-за омонимии флексии *-ая* не распознались деепричастия в парадигме ПРОЧИТ-ать. Можно надеяться, что переход к минимальному интервалу позволит правильно квалифицировать многие из этих форм.

2.2.5. Уточнение парадигм

Примем достигнутый результат как некоторую реальность и будем постоянно возвращаться к нему при дальнейшем анализе.

Снимая ранее введенные пороги, попытаемся уточнить состав обнаруженных парадигм. Начнем с того, что допустим аффиксы длиной более четырех букв. Теперь потенциальным аффиксом может стать, например, *-анного* в слове *воспитанного*, а у основы *воспит-* открывается 52 продолжения, в том числе такие длинные как *воспитательного*¹³. Аффиксы *-ал, -ала, -али, -ался, -ать, -ывал* принадлежат парадигме 3 (ПРОЧИТ-ать), значит наша основа принадлежит той же парадигме¹⁴. Аффиксы *-ание, -анию, -ания* относятся к парадигме 40 (для которой ярлыком выбрана рассматриваемая основа), теперь к трем стандартным аффиксам добавляются *-анием, -ании, -анию, -анье*; одновременно доказана производность субстантивов на *-ание* от глаголов. Аффиксы *-ателей, -ателем, -атели, -атель* естественным образом подключаются к парадигме 26 (СОЧИНИ-тель), у получившейся основы *воспитатель* есть три продолжения: *-ного, -ные, -ный*, где *-н-* интерпретируется как отыменной суффикс, а *-ого, -ые, -ый* относят эти слова к парадигме 17 (ТЯЖЕЛ-ый).

Остается еще 22 продолжения, все они начинаются на *-ан*, из них 19 содержат еще одно *-н-*, а в трех случаях находим аффиксы *-о, -а, -ы*. Набор этих трех аффиксов подобен парадигме 7 (ОПРЕДЕЛ-ен), только на этот раз парадигма 7а разрастается за счет суффикса *-н-* и адъективных аффиксов (из парадигмы 17) *-ая, -ого, -ой, -ые, -ый, -ых*. Таким образом, достраивается парадигма пассивного причастия (прилагательного). Многочисленные факты показывают, что аналогично расширяется парадигма 7, но там предпочтительной оказывается трактовка адъективов как прилагательных, причастия (вроде *полученный*) встречаются реже. У основы *воспитанн-* находим аффиксы *-ик, -ика, -икам, -ики, -иков, -иком, -ику* (все из парадигмы 23) и *-иц, -ица, -ице, -ицей, -ицу, -ицы* (ср. парадигму 29).

Если бы вместо *воспит-* рассматривалась основа *взволн-*, сразу была бы установлена ее принадлежность к парадигме 12, к парадигме 7а была бы отнесена основа *взволнованн-*, но мы не нашли бы ничего похожего на парадигмы 23, 26, 29, 40.

У основы *выдум-* сразу устанавливается принадлежность к глагольной парадигме 3 (включая производную основу *выдумыв-*), обнаруживается основа *выдуманн-* (парадигма 7а), а также продолжения *-ка, -ке, -ки, -кой, -ку, -ок* (парадигма 32). Здесь же находим слово *выдумщик*. Как протоаффикс *-щик* встретился в нашем материале 21 раз, показал сильную связь с *-шки* S=71 (основы *барабан-, бунтов-, караул-, контор-, миллион-, ростов-*), что указывает на парадигму 23, но своей явной парадигмы не создал. Суффикс *-щик* с вариантами *-ыщик* после *л* (*могил-, караул-, носил-, парил-*) и *-овщик* после сочетаний согласных (*гурт-*) с семантической точки зрения очень ясен — все это обозначения людей, главным образом, по роду деятельности, ср. *артель-, заговор-, камен-, натур-, обой-, откуп-, псалом-* и т. д.¹⁵

Перечень основ, данный выше при постулировании парадигмы ЗНАКОМ-ство, включал и такие, которые кончались на *е* (*ничтоже-, товарище-, человеке-*).

¹³ Напомним, что речь идет о графических словах, встретившихся более 7 раз. Во всем исследуемом корпусе у левой цепочки *воспит-* находим 162 продолжения.

¹⁴ Основа *воспитыва-* имеет продолжения *-ет, -ется, -л, -ла, -лась, -ли, -лись, -лся, -ть, -ться, -ют*, явно указывающие на парадигму 1 (НАЗЫВА-ть); при этом показатель *-ыва-* интерпретируется как суффикс, образующий глаголы от глаголов.

¹⁵ Внутренняя форма иногда затемнена, ср. *денщик, кормщик, прапорщик, ямщик*.

Разрешив пятибуквенные аффиксы, проверим на «парадигматичность» конечную цепочку -ество. Выяснилось, что перед такой цепочкой появляются ж – муж-, невеж-, ханж-, худож- (также ж<г: досуж-, множ-, убож-, торж-); ч – зодч-, колич- (также ч<к: велич-, владыч-, высоч-, дурач-, заступнич-, одиноч-, отроч-, сопернич-, сотруди́нич-, чиновнич-, чудач-; ц<к: молодч-; ш – излиш-; щ – изящ-, имущ-, могущ-, общ-, сущ-.

Такое же чередование повторяется и перед цепочкой -еский: ж<г: бож-, друж-, супруж- (также ж<з: княж-); ч – ребяч-, электрич- (также ч<к: иноч-, католич-, мошеннич-, мученич-, пророч-, челове́ч-; ч<ц: девич-, купеч-, младенч-, отец-, старч-, творч-); ш – юнош- (также ш<х: монаш-); щ – вещ-.

Продолжая движение от конца графического слова к его началу, мы можем подойти к коротким основам (менее четырех букв). Так, слово *взяточничество* производно от *взяточник*, которое производно от *взятка* (ср. *будочник* < *будка*, *булочник* < *булка*, *лавочник* < *лавка*, *лодочник* < *лодка*, *сказочник* < *сказка*). Последнее слово прямо ведет нас к четырехбуквенной основе *взят-* (парадигмы 30 и 38). Еще один шаг, и мы окажемся перед трехбуквенной основой *взя-*, сочетающимся с характерными аффиксами парадигмы 1: -в, -вши, -вший, -л, -ла, -лась, -ли, -лись, -ло, -лось, -лся, -ть, -ться. Интересно, что аффиксы -ем, -емся, -ет, -ете, -ется, -ешь, ожидаемые в парадигме 1, здесь отсутствуют. С другой стороны, существует протопарадигма *возьм-* /-ем, -ет, -ете, -ешь, -и, -ись, -ите, -у, -усь, -ут/. Такие дефектные варианты парадигмы НАЗЫВА-ть (с отсутствующими формами прошедшего времени) наблюдаем у четырех десятков глаголов, где основы настоящего и прошедшего времени существенно различаются, ср. примеры трехбуквенных основ (корней?) с их префиксами:

каж- + до-, за-, на-, о-, от-, по-, при-, рас-, с-, у-;
пиш- + за-, на-;
бер- + вы-, раз-, со-, у-;
зов- + вы-, на-, по-;
войд- + вы-, до-, за-, на-, обо-, ото-, подо- по-, про-, со-, у-;
прочт- + со-.

Очень короткая (двубуквенная!) основа *ям-* показывает, с какими проблемами может столкнуться ДСАТ при отказе от ограничений длины основы. В нашем материале такое начало засвидетельствовано у 26 слов:

ям	яме	ямочки	ямскую	ямщиками
яма	ямки	ямская	яму	ямщики
яман	ямку	ямские	ямщик	ямщиков
яманов	ямой	ямских	ямщика	ямщиком
ямах	ямочками	ямской	ямщикам	ямщику
				ямы

1) Основа *ямщик-* /-0, -а, -ам, -ами, -и, -ов, -ом, -у/ несомненно относится к парадигме 23.

2) Основа *ямск-* /-ая, -ие, -их, -ой, -ую/ принадлежит парадигме 20.

3) Слова *ям, яма, ямах, яме, ямой, ямой, ямы* соответствуют парадигме 27.

Производная отсюда основа *ямк-* относится к парадигме 30. От этой основы произведена основа *ямочк-*, такая деривация подтверждается многими десятками пар типа

бабочка	бабка ¹⁶	записочка	записка	курточка	куртка
бутылочка	бутылка	кибиточку	кибитк	лавочка	лавка
водочки	водка	колясочка	коляска	лодочка	лодка
девочка	девка ¹	коробочка	коробка	ложечки	ложка
деточки	детки	корочки	корка	маточка	матка
дырочку	дырка	красоточка	красотка	минуточку	минутку

4) Остатятся два слова *яман* и *яманов*¹⁷.

¹⁶ Заметим, что семантические различия в этой паре не улавливаются ДСАТ в микроинтервале.

¹⁷ Обращаясь к низкочастотным словам, найдем там слово *яманова* – напрашивается интерпретация посредством парадигмы 41 (фамилии). Два слова *яманный* и *яманской*

В какой мере эти четыре случая семантически связаны друг с другом? Ответ на этот вопрос даст переход к среднему интервалу.

На пути от ядра морфологии к коротким протоосновам появляются новые парадигмы или варианты существующих парадигм. Рассмотрим в качестве примера парадигму 18 (ХОРОШ-ий) с 13 аффиксами /-ая, -его, -ее, -ей, -ем, -ему, -ею, -ие, -ий, -им, -ими, -их, -ую /, из которых три аффикса (-ая, -ее, -ую) представлены в парадигме 17 (ТЯЖЕЛ-). Пусть набор шести из оставшихся десяти аффиксов будет достаточным для того чтобы считать какую-то короткую основу принадлежащей к нашему типу. Оказывается, что набор восьми аффиксов /-его, -ей, -ем, -ему, -ею, -им, -ими, -их¹⁸/ зафиксирован при основах мо- (f=49741), сво- (f=47940), тво- (f=9600) и с- (f=6470). У первых двух основ к этому набору добавляются аффиксы -е, -и, -ю, -я, а основа с- добавляет -е, -ию, -ия, -ю. Учитывая громадную частоту, можно постулировать особую парадигму для мо- и тво-; с несколько меньшим основанием то же можно сказать и об основе с-.

Семь аффиксов находим у основы вс- (f=136002) /-е, -его, -ей, -ему, -ею, -ю, -я/, к этому набору присоединяется и аффикс -ех.

Две основы ваш- (f=22990), наш- (f=21070) имеют семь аффиксов, общих с парадигмой 17; они добавляют -о, -а, -и, -у, создавая особую парадигму.

Уникальную парадигму образует основа эт- (f=189906) /-а, -и, -им, -ими, -их, -о, -ого, -ой, -ом, -ому, -ою, -у/. Этот набор очень напоминает парадигму 20; переход к минимальному интервалу позволит решить, относится ли сюда слово этот (f=15204).

2.2.6. Префиксы

Алгоритм формирования парадигм при дистрибутивном анализе правых концов графических слов оказался успешным: выделено несколько десятков парадигм, группирующихся в три больших класса. Успех этот объясняется самим устройством русской морфологии – именно в конце слова сосредоточены все словоизменительные и многие словообразовательные аффиксы.

Переход к анализу начала графического слова ставит перед исследователем ряд сложных проблем. Начало ранжированного списка пар протопрефиксов кажется странным:

		S			S
0	не	336	дру	кня	250
з	н	316	неп	п	242
еди	таи	302	0	по	202
з	нез	289	ег	фед	201

Пара /0- – не-/ образует узкую, но чрезвычайно мощную парадигму (1022 основы), ср. *аккуратно, богатого, возможность, законный, искренность, натуральным, обыкновенные, подвижное, расположение, складные, уважения, человеческим* и т. п. Эта парадигма не показывает никаких значимых связей с какими-либо другими префиксами¹⁹. Поведение не- аналогично поведению ся/сь в суффиксальных парадигмах.

Но дальше начинаются трудности. Пара /з- – н-/ представлена в 37 парадигмах с 420 основами, но почти 90% всех основ сосредоточено в эксклюзивной парадигме, образующиеся при этом основы неизменно начинаются на букву а (*-абегал, -адумала, -анимается* и т. п.). Это заставляет думать о двух префиксах – за- и на-. В третьей строке списка находим пару *еди- – таи-*, где основы неизменно начинаются на -н (*-нственной* и т. п.). Ясно, что здесь представлены основы *един-* и *таин-* с характерными продолжениями адъективной парадигмы, отнюдь не префиксы. Основами

выступают скорее как свидетельство интерпретации слова *яман* как представителя парадигмы 21.

¹⁸ Любопытно, что все эти семь аффиксов (+ -ею) выступают и как графические слова.

¹⁹ Остаются статистически незначимыми редкие пары: (не) *возможными/всевозможными*, (не) *сказанное/высказанное*, (не) *вольного/довольного*, (не) *благополучно/преблагополучно*, (не) *образования/преобразования*.

являются *дру-* и *кня-* (с продолжениями *-жеское*, *-зьями* и т. п.) и *ег-* и *фед-* (с неизменными продолжениями на *-ор*). От проникновения основ в список префиксов можно в какой-то мере застраховаться, вводя пороги. Потребуем, чтобы кандидат в префиксы имел более 20 основ и встречался, по крайней мере в 11 протопарадигмах²⁰.

Далее следует случай, очень характерный именно для начала графического слова, — один член пары повторяется как хвост другого члена. Таковы *з-/нез-*, *неп-/п-*, *д-/нед-* (S=176), *доз-/поз-* (S=162), *зах-/х-* (S=113), *з-/поз-* (S=106). Вариант той же ситуации — совпадение хвостов обоих членов: *изъ-/объ-* (S=127), *исп-/нап-* (S=115). Исключим такие пары из ранжированного списка, который примет теперь следующий вид:

336	0	не	80	0	про	69	0	пре	59	от	у
202	0	по	79	0	на	68	от	по	59	при	с
125	0	за	79	вы	за	67	за	с	58	до	за
91	0	у	78	при	у	66	вы	у	58	за	о
91	за	при	76	0	рас	66	по	про	57	0	без
89	0	при	75	0	о	65	на	по	57	вы	с
89	за	по	75	по	с	64	вы	от	57	от	при
88	зна	му	74	вы	при	62	0	со	56	на	от
87	по	у	73	за	у	62	вы	на	56	по	рас
87	0	с	73	на	при	62	от	раз	56	с	у
86	0	вы	73	от	с	61	за	от	55	0	от
85	вы	по	73	по	при	60	вы	до	55	за	под
82	0	бес	72	за	на	60	на	у	52	раз	с
81	вы	про	72	за	про	59	вы	под	52	скр	заб

В этом списке особняком стоят две пары: *0-* — *без-* и *0-* — *бес-*. Обе пары создают мощные эксклюзивные парадигмы */0-* — *без-/* с 67 основами (*-бедно*, *-вредно*, *-жалостный*, *-звучно*, *-мятежный*, *-нравственная*, *-образную*, *-различно*, *-умно* и др.) и */0-* — *бес-/* с 97 основами (*-конечная*, *-полезно*, *-порядка*, *-примерно*, *-сильным*, *-смертных*, *-сонные*, *-толково*, *-характерный*, *-цветной*, *-честные*, *-шумно*). В обеих парадигмах 90% составляют адъективные основы; перед конечной буквой у них, как правило, наблюдается *н*.

Буквенное и дистрибутивное сходство двух аффиксов заставляет предположить, что это (орфографические) варианты одного префикса. Ответ ясен с позиций семантической лингвистики — после *бес-* все основы начинаются на глухой согласный; в рамках ДСА об этом же свидетельствует таблица 2.4. Это наблюдение позволяет объединить в одном префиксе также *из-/ис-*, *раз-/рас-*, *воз-/вос-*.

Особняком стоит префикс *пре-*, значимо связанный только с *0-* (S=69) и с *по-* (S=43), он сочетается с 24 основами: *-доставлена/а/ы*, *-доставляет/ли*, *-зрительную*, *-имуществом*, *-исполненный*, *-образованием*, *-освященный*, *-подавались*, *-следования*, *-следователей*, *-увеличение/ия*, *-увеличивает/л/ть*, *-хорошенький/ая/ие/ой*.

По мере продвижения вниз по списку открывается исключительно широкая парадигма. Внутренние связи ядра этой префиксальной парадигмы показаны в следующей матрице:

	0	в	вы	до	за	ИЗ	на	о	об	от	по	под	при	про	РАЗ	с	у
0		28	86	47	125	35	79	75	26	55	202	36	89	80	76	87	91
в	28		36	27	31	14	28	12	34	32	29	33	34	24	12	30	24
вы	86	36		60	79	24	62	46	28	64	85	59	74	81	44	57	66
до	47	27	60		58	16	40	46	10	45	51	28	48	49	27	37	51
за	125	31	79	58		28	72	58	25	61	89	55	91	72	50	67	73
ИЗ	35	14	24	16	28		46	16	30	20	26	10	33	19	28	24	22
на	79	28	62	40	72	46		44	25	56	65	34	73	41	38	47	60
о	75	12	46	46	58	16	44		10	24	47	27	32	26	76	34	39
об	26	34	28	10	25	30	25	10		33	27	30	26	15	39	36	14

²⁰ Тем не менее, в ранжированный список проникли *зна-* и *му-* (S=88; продолжения — *-чительная* и т. п.), *заб-* и *скр-* (S=52; продолжения — *-ывать* и т. п.) *р-* и *ут-* (S=41; продолжения — *-ешать* и т. п.).

от	55	32	64	45	61	20	56	24	33		68	51	57	40	62	73	59
по	202	29	85	51	89	26	65	47	27	68		34	73	66	56	75	87
под	36	33	59	28	55	10	34	27	30	51	34		47	28	28	39	37
при	89	34	74	48	91	33	73	32	26	57	73	47		42	35	59	78
про	80	24	81	49	72	19	41	26	15	40	66	28	42		31	33	42
РАЗ	76	12	44	27	50	28	38	76	39	62	56	28	35	31		52	40
с	87	30	57	37	67	24	47	34	36	73	75	39	59	33	52		56
у	91	24	66	51	73	22	60	39	14	59	87	37	78	42	40	56	

Приведем несколько примеров сочетания основ и префиксов:

вести	0	в	воз	вы	до	за	из	на	об	от	по	под	при	про	раз	с	со (!)	у
писал	0	в	вы	до	за	ис	на	над	о	от	по	под	при	про	рас	с		
ходил	0	в	вс	вы	до	за	ис	на	об	от	по	под	при	про	с	у		
ведет	0	в	вы	до	за	из	на	от	по	под	при	про	раз	с	у			
водит	0	в	вы	до	за	на	об	от	по	под	при	про	раз	с	у			
летел	0	в	вы	до	за	на	об	от	по	под	при	про	с	у				
ставил	0	в	вы	до	за	на	о	от	по	под	при	рас	со	у				
пуская	0	в	за	вы	до	ис	о	от	под	про	рас	с	рас	у				
тянул	0	в	вы	за	на	по	под	при	по	про	с							
бегает	0	в	вы	за	из	от	под	при	про	с	у							
садил	0	в	вы	за	на	о	по	про	с	у								
тащили	0	в	вы	от	по	при	рас	с	у									

Громадное большинство основ относится к глагольной парадигме. Поэтому префиксы, приведенные в матрице, можно назвать превербами. Все они совпадают с отдельными графическими словами, которые, как выяснится на минимальном интервале, являются предлогами (за исключением *раз*).

Превербами можно назвать и некоторые префиксы, не вошедшие в ядро²¹. Прежде всего, следует упомянуть варианты префиксов с исходом на *о* — *во-* (3), *изо-* (5), *обо-* (6), *ото-* (6), *со-* (23). Лишь *со-* сильно связан с ядром, но внутренние связи этой маленькой парадигмы достаточно велики (среднее $S=21$), а это явное указание на какие-то особые дистрибутивные свойства, отличающие ее от ядра парадигмы превербов.

Примерами служат

гнать	0	во	вы	до	за	из	на	ото	про	со	у
рвать	0	вы	изо	на	обо	ото	по	пре	про	со	у
брать	0	вы	за	из	на	обо	ото	при	про	со	у

Другие дистрибутивные свойства обнаруживаются у превербов с исходом на *ъ*: *въ-* (6), *изъ-* (1), *объ-* (6), *отъ-* (4), *съ-* (3), со средним значением $S=50$. В отличие от вариантов на *о* варианты на *ъ* обнаруживаются только внутри графического слова.

Примерами служат

ехали	0	въ	вы	до	за	на	объ	отъ	по	при	про	съ	у
езжал	0	въ	вы	за	на	объ	отъ	при	про	съ	у		
ездил	0	изъ	объ	съ									
являет	0	за	изъ	объ									
яснялся		изъ	объ										

Особыми превербами оказываются *воз-* (13), *вос-* (8), *вз-* (14), *вс-* (10), и, может быть, *над-* (1):

носиться	0	воз	до	от	с	гласил	0	воз	о	при	
кликнуть	0	вос				торжествовать	0	вос			
дохнуть	0	в	вз	из	от	думают	0	вз	вы	по	при
крикнет	0	вс				тряхнет	0	вс			
лежало	0	над				ругается	0	над			
писано	0	за	на	над	о	под	про	рас			

²¹ В скобках указано среднее значение S связи данного префикса с ядром парадигмы. Отброшенные на предыдущем шаге алгоритма (как совпадающие по своему исходу) теперь они «реабилитированы» как варианты превербов.

2.2.7. Парадигмы и размер исследуемого корпуса

Результаты дистрибутивно-статистического анализа в большой мере зависят от размера изучаемого корпуса текстов. В какой степени этот фактор влияет на процесс открытия морфологии? Чтобы ответить на этот вопрос алгоритм, описанный в пункте 2.2, был применен к тексту «Войны и мира» (F=450000) и к текстам «Независимой газеты» 1996–2000 гг., служившим материалом для пункта 2.1.

Роман Толстого входит в наш корпус прозы, составляя 3% его объема. При меньшем объеме корпуса можно ожидать и меньшие значения S. Объем корпуса «Независимой газеты» в два раза больше корпуса прозы, значит, и значения S должны быть больше. Сравним начало ранжированного списка пар протоаффиксов в этих трех корпусах:

Корпус прозы			«Война и мир»			«Независимая газета»		
1695	ми	х	382	ми	х	2721	ми	х
1069	ась	ся	235	м	х	1719	м	х
1024	м	х	210	м	ми	1644	го	й
842	ась	ись	173	й	м	1555	м	ми
792	м	ми	156	ась	ся	1531	й	х
766	ись	ся	142	й	х	1427	ась	ось
758	ась	ось	130	ись	ся	1366	го	м
730	й	х	129	е	й0	1260	й	м
626	й	м	125	е	х	1213	ась	ись
589	ись	ось	123	го	й	1154	ись	ось
585	го	му	121	ась	ись	1110	е	й
583	е	х	121	0	сь	1091	ись	ся
537	го	й	108	0	ми	1071	е	х
526	0	ся	105	го	му	1066	ась	ся
517	ось	ся	104	0	и	907	го	м

Однако при этих различиях в значениях S общая картина агрегированных парадигм остается все той же. Особенно близкими они кажутся при сравнении Корпуса прозы и «Войны и мира».

Сравнение Корпуса прозы и «Независимой газеты» показывает некоторое различие в последовательности формирования парадигм, хотя окончательный набор парадигм, в основном, тот же. Впрочем, отдельные расхождения крайне интересны.

Среди диахронических различий самой яркой является отсутствие аффиксов -ою и -ею в парадигмах «Независимой газеты». Полностью исчез и протоаффикс -с (словоерс). Среди протоаффиксов в НГ не обнаружен -ши. В Корпусе прозы связи этого протоаффикса явно указывают на омонимию.

Во-первых, -ши связан с -0 (S=32), -ша (60), -ше (44), -шу (43), ср. основы адмираль-, бригадир-, генераль-, губернатор-, доктор-, камергер-, капитан-, матре-, павлу-, петру-, помпадур-, порфи-, почтмейстер-, предводитель-, торго- (!).

Во-вторых, -ши связан с -шись (S=49) и с такими характерными глагольными аффиксами, как /-ает, -ай, -ал, -ала, -али, -ать, -аю/, он сочетается с основами на в (встав-, выпив-, забыв- и т. п.), но иногда и с другим исходом вырос-, вышед-, подпер-, привык-, пришед-. Вряд ли можно свести исчезновение -ши в НГ исключительно к диахронии. Здесь на первый план выступают жанровые различия.

Среди глагольных аффиксов парадигмы 3 можно заметить исчезновение -айте и -аете, а также некоторое ослабление S у аффиксов -ай, -аешь, что коррелирует с уменьшением доли диалогов в газете. С другой стороны, можно отметить рост S у аффикса -аем и у серии аффиксов -ан, -ана, -ано, -аны, коррелирующей с характерной для газеты большей долей пассивных конструкций. Тем не менее, парадигма 3 как целое хорошо сохраняется и в «Независимой газете»: из 22 аффиксов исчезли лишь два.

Разрастается и укрепляется парадигма 12 ЧУВСТВ-овать; теперь в ней 7 аффиксов /-овав, -овал, -ован, -уем, -ует, -уют, -уя/. Уже среди протопарадигм обнаружен набор 4 аффиксов с 28 основами: агитир-, акцентир-, апеллир-, аргументир-, декларир-, демонстрир-, дискредитир-, игнорир-, иллюстрир-, исслед-, комментир-, координир-, лоббир-, оперир-, провоцир-, пропагандир-, реагир-, регулир-, рекламир-, символизир-, способств-, стимулир-, фиксир-,

финансир-, форсир-, характериз-, эксплуатир-, экспортир-. В агрегированной парадигме все семь аффиксов имеют основы гарантир-, использ-, организ-; шесть аффиксов имеют основы блокир-, констатиру-, ликвидир-, проголос-, ратифицир-, реализ-, рекоменд-; по пять аффиксов имеют 26 основ; число основ с четырьмя аффиксами превышает шестьдесят.

Некоторый урон потерпела и парадигма 33 («отчества») из набора восьми аффиксов ушли -вне и -вну. Дательный и винительный падежи отчеств женщин в газете встречаются реже, чем в прозе.

Новая парадигма 44 (МУРМАН-ск) характеризуется тремя аффиксами /-ск, -ска, -ске/, ср. каспий-, ноябрь-, рыбин-, свердлов-, хабаров-, ульянов-, черкес-.

Значительно пополнились субстантивные парадигмы с интернациональными аффиксами. Так, парадигма 41 имеет теперь 8 аффиксов /-тор, -тора, -торы, -цией, -ции, -ций, -цию, -ция/. Все восемь аффиксов засвидетельствованы у основ административ-, инспек-, инструк-, компози-, конструк-, организа-. По 7 аффиксов имеют интерпрета-, информа-, коллек-, консерва-, координа-, приватиза-, провока-, реставра-, по 6 – генера-, импровиза-, интегра-, ликвида-, реформа-, по 5 – дирек-, опера-, регистра-, редак-. От этой парадигмы производна парадигма 41а ПУБЛИК-ация /-ации, -аций, -ацию, -ация/ с основами демонстр-, концентр-, манифест-, организ-, ориент-, публик-, рекоменд-, по 3 аффикса имеют аргумент-, аттест-, дискредит-, документ-, индекс-, конверт-, консерв-, констат-, координ-, ликвид-, презент-, профан-, регистр-, регламент-, резерв-, эксплуат-.

Здесь же можно заметить парадигму 41б ОРГАНИЗ-ация, где к указанным 4 аффиксам присоединяются /-ма, -ме, -мом/. Все 27 основ кончаются на з: абсолютизм-, бюрократизм-, глобализм-, гуманизм-, демократизм-, идеализм-, индивидуализм-, исламизм-, капитализм-, коллективизм-, либерализм-, механизм-, милитаризм-, модернизм-, монополизм-, национализм-, натурализм-, оптимизм-, радикализм-, рационализм-, реализм-, регионализм-, социализм-, федерализм-, формализм-, централизм-.

Отсюда один шаг до парадигмы 45 КАПИТАЛ-изм с семью аффиксами /-изм, -изма, -изме, -изму, -ист, -иста, -исты/. По 7 аффиксов имеют основы коммунал-, национал-, постмодерн-, социал-, террор-, по 6 – монопол-, по 5 – авангард-, авантюра-, империал-, сепарат-, федерал-, по 4 – авторитар-, антисемит-, гомосексуал-, интернационал-, консерват-, либерал-, модерн-, патриот-, профессионал-, радикал-, тоталитар-, фундаментал-.

В Корпусе прозы была зафиксирована пара -нт/-нта с основами ареста-, дилета-, музыка-, официант-, проце-, секунда-, сержант- (!), фабрика-. В "Независимой газете" зафиксирована уже агрегированная парадигма 46 МУЗЫКА с 4 аффиксами /-нт, -нта, -нтов, -нты/ со средним S=126, ср. дебюта-, консульта-, номина-, спекуля-, фигура-, эмигра-; по 3 аффикса у основ вариация-, демонстра-, иммигра-, официант-, протеста-, проце-, сержант-, фабрика-.

Наконец, в «Независимой газете» находим четырехчленную парадигму 47 КУЛЬТУРА /-ура, -уре, -уру, -уры/ со средним S=165, ср. аппарат-, архитектор-, диктатор-, кандидат-, литератор-, режиссер-, скульптор-. Три аффикса сочетаются с основой агент-, адвокат-, аспирант-, префект-, резидент-.

В корпусе «Независимой газеты» сохраняются все префиксы, полученные на Корпусе прозы. К ним добавляются:

1) Два интернациональных префикса де- и ре-, ср. -формации, -конструкции, -монтировать. Только с де- сочетаются -долларизации, -колонизации, -легат, -марша, -милитаризация, -мобилизации, -монополизация, -монтаж, -номинация, -популяции, -приватизации, -русификации, -стабилизация, -централизация и др.²² С ре- сочетаются -активный, -акционеров, -анимация, -генерация, -организовать, -презентации, -трансляции, -формам и др.

2) Два компонента-аббревиатуры: мин- (основы -атома, -госимущества, -здоровом, -культ, -науки, -обороны, -печати, -природы, -связи, -содружества, -соцзащиты, -транс, -экономики) и более мощный гос- (126 основ, включая -автоинспекции, -аппарат, -безопасность, -бумаги, -бюджет, -гарантий, -границе, -департамент, -долга, -дотаций, -думой, -заказ, -казну, -канцелярии, -капитализм, -комиссию, -компания, -контроль, -концерта, -министр, -монополия, -облигации, -органы, -оркестры, -пакет, -переворот, -плане, -подам (!), -поддержке, -посредников, -предприятие, -премия, -программы-, расходов,

²² Уровень ошибок (типа -нежный, -ржавой, -санта) достигает здесь 20%.

-санэпиднадзор, -секретарь, -сектор, -служащие, -служба, -собрание, -собственность, -советник, -советом, -средств, -строительства, -структурах, -тайны, -телевидения, -университета, -управление, -учреждениях, -фонда, -чиновник, -языка).

3) Два префикса, сочетающиеся с адъективными основами: меж- (78 основ, например, -афганского, -банковский, -ведомственный, -клановой, -корейских, -личностных, -муниципальный, -национальное, -отраслевой, -парламентский, -районного, -региональный, -религиозной, -таджикский, -фракционной, -этнический) и все- (96 основ: -белорусского, -великое, -китайского, -кубанского, -мирный, -могущий, -народный, -непременно, -общий, -российский, -сильных, -союзный, -татарский, -украинский).

2.3. Английский язык

Процедура, описанная в п. 2.2, применяется здесь к корпусу текстов романов Диккенса (F = 4080 тыс. словоупотреблений).

Как и в случае русского языка начнем с формирования ранжированного списка значений S для пар протоаффиксов.

444	0	s	207	ies	y	155	d	rs	129	man	men
366	'	g	206	d	r	149	e	es	124	five	four
277	on	ons	199	r	rs	148	ion	ions	123	t	tly
265	ed	ing	183	ng	ons	145	ied	ying	121	g	gly
255	ng	on	178	ly	ness	144	r	st	121	0	ed
220	0	's	168	ve	vely	143	0	d	121	e	ing
212	0	ly	160	on	ve	139	's	s	119	ons	ve
208	g	gs	158	d	s	138	'	gs	117	0	ing

Прежде всего, в этом списке поражает большое число пар протоаффиксов, начинающих на одну и ту же букву (или цепочку букв): on/ons, g/gs, r/rs, ve/vely, e/es, ion/ions, man/men, five/four. Временно исключим подобные пары из нижеследующего более пространныго списка.

444	0	s	158	d	s	116	five	two	96	ing	s
366	'	g	155	d	rs	116	four	two	95	five	six
265	ed	ing	145	ied	ying	115	ce	t	95	four	six
255	ng	on	144	r	st	115	on	vely	91	ed	ion
220	0	's	143	0	d	114	es	ing	90	'	0
212	0	ly	139	's	s	106	'	gly	89	cy	tly
207	ies	y	138	'	gs	105	five	nine	89	e	ion
206	d	r	121	0	ed	105	four	nine	89	six	two
183	ng	ons	121	e	ing	99	-six	-two	86	0	ness
178	ly	ness	119	ons	ve	98	nine	two	84	ce	tly
160	on	ve	117	0	ing	97	ed	s	83	's	s'

Многие из этих пар совпадают с эксклюзивными протопарадигмами, иногда очень мощными. Так, протопарадигма /0 – s/ включает 2690 основ, протопарадигма /' – g/ имеет 126 основ (boxin-, breedin-, carryin- и т. п., все с исходом in), протопарадигма /ed – ing/ имеет 188 основ, протопарадигма /ng – on/ видна в 108 основах (substituti-, suffusi-, supervisi-, все с исходом на i), в протопарадигме /0 – 's/ 681 основа, протопарадигма /0 – ly/ имеет 730 основ, протопарадигма /ies – y/ имеет 176 основ, протопарадигма /d – r/ имеет 117 основ (lurke-, mistruste-, opene-, все с исходом на -e). В протопарадигме /d – s/ 190 основ, из которых три основы – с исходом на ', а все остальные – с исходом на -e). Протопарадигма /ies – y/ имеет 176 основ, протопарадигма /r – st/ имеет 90 основ (dulle-, earlie-, easie-, faire-, все с исходом на -e); протопарадигма /0 – d/ имеет 326 основ, из которых одна – с исходом на ', а все остальные – с исходом на -e). В протопарадигме /0 – ing/ 148 основ; в протопарадигме /0 – d/ 88 основ; в протопарадигме /0 – ness/ 70 основ. Перечисленные 14 протопарадигм могут стать агрегированными парадигмами или их частями. Их совокупная мощность превышает 6000 основ (около 40%) всех основ. Контраст с началом русского списка налицо.

Обратимся ко второй строке нашего списка, которая должна вызвать недоумение у человека, знающего английский язык. Соотношение '/g, а точнее – n'/ng объясняется стремлением Диккенса показать лондонское простонародное произношение [n] вместо литературного -ng. В корпусе романов Диккенса встретилось 227 основ, допускающих и ' и ng, а точнее – in/ing (типа givin' и giving)²³.

При всей бедности английской морфологии довольно скоро выявляется четырехчленная агрегированная парадигма SAILOR /-0, -'s, -s, -s'/ со средним значением S=163; как эксклюзивная парадигма она представлена у 49 основ (banker, lawyer, lover, merchant, native, physician, reader, sheriff, visitor и т. п.). Еще больше основ (342) у трехчленного варианта /-0, -'s, -s/. Другой трехчленный вариант /-0, -s, -s'/ охватывает 23 основы (counsellor-, courtier-, dagger-, million-, postilion-, seller-, shilling-, worker- и др.). Таким образом, сформировалась главная субстантивная парадигма. Лишь минимальный интервал покажет, что к этой же парадигме относится протопарадигма /0 – s/ с ее 2690 основами.

Вокруг протопарадигмы /-ed, -ing/ образуется главная глагольная парадигма /-e, -ed, -es, -ing/ со средним значением S=149. С той же протопарадигмой связан протоаффикс -0 (S=119). Следующая матрица показывает внутреннюю неоднородность агрегированной парадигмы²⁴:

	0	ed	ing	e	es	ion	er	ive
0		121	117	3	43	16	42	8
ed	121		265	111	114	91	60	29
ing	117	265		121	114	76	77	25
e	3	111	121		149	89	27	27
es	43	114	114	149		45	25	20
ion	16	91	76	89	45		-	72
er	42	60	77	27	25	-		-
ive	8	29	25	27	20	72	-	

Как видим, внутри агрегированной парадигмы проглядывают два подмножества.

Эксклюзивная парадигма CONDEMN /-0, -ed, -ing/ охватывает 145 основ: abolish-, disarm-, invert-, lessen-, roast-, undress- и т. п. С аффиксами -ed и -ing связан также аффикс -s (S=96). Эксклюзивная парадигма /-0, -ed, -ing, -s/ включает 236 основ: allow-, cheat-, demand-, enter-, greet-, kneel-, pluck-, result- и др.

В эксклюзивной парадигме DECID-e /-e, -ed, -es, -ing/ находим 153 основы, ср. achiev-, decid-, ensur-, mingl-, requir- и др.

К глагольной парадигме примыкают кластеры

/-able, -ed, -ing/

с основами comfort, consider, deplor-, detest, favour, honour, reason, remark, respect, unquestion; и

/-ed, -ing, -ment, -s/

с основами abandon, adjust, appoint, assail, assign, attain, bewilder, command, conceal, consign, content, depart (!), develop, disappoint, employ, enlighten, entertain, environ, govern, ill-treat, invest, present, resent, sustain, wonder.

На самой периферии глагольной парадигмы находим набор /-ed, -ion, -or, -or's, -ors/, в котором -or создает имена деятеля подобно намного более активному -er. Сюда входят 35 основ: abett-, audit-, calculat-, calumniat-, collect, conduct, conjur-, conquer, contract, counsell-, credit, direct, execut-, exhibit, govern, imitat-, inherit, inspect, instruct, narrat-, oppress, originat-, perpetrat-, persecut-, possess, profess, project, prosecut-, protect, solicit, speculat-, survey, surviv-, torment, visit.

²³ В "Войне и мире" встречается пара г'/р, передача Толстым картавости Денисова.

²⁴ К пяти несомненным аффиксам добавлены -ion, -er и -ive. Первый аффикс показывает максимальную связь с -ions (S=148), второй с -ers (S=99) и с -er's (S=48), что доказывает их субстантивный и словообразовательный характер. Оба аффикса очень продуктивны (более 300 основ у каждого из них). Словообразовательный аффикс -ive образует прилагательные, максимальную связь аффикс -ve имеет с -vely (S=168).

Кластер адъективной парадигмы включает аффиксы /-0, -er, -est, -ly, -ness/. Внутренние связи даны в следующей матрице.

	0	ly	ness	er	est
0		212	86	42	25
ly	212		178	33	79
ness	86	178		20	58
er	42	33	20		56
est	25	79	58	56	

Как эксклюзивная парадигма CLEVER включает также blessed-, crooked-, forlorn-, shrewd-, wretched-. С пятью аффиксами сочетаются также bitter-, bright-, narrow-, pleasant-, prompt-, strait-, strict-, tender-. По 4 аффикса имеют abrupt-, awkward-, broad-, cheerful-, clear-, distinct-, helpless-, honest-, loving-, plain-, quaint-, ragged-, shrill-, slight-, wicked-, willing-.

Протопарадигма /-0, -ly, -ness/ имеет 101 основу: childish-, earnest-, impressive-, earnest-, skilful-, useless- и т. п. Но самой мощной оказалась протопарадигма /-0, -ly/, охватывающая 730 основ: anxious-, doubtful-, fifth-, infernal-, mocking-, political- и т. п.

К адъективной парадигме примыкает пара -ity – -ly (S=34), в которой аффикс -ity создает деадъективные субстантивы, ср. absurd, angular, benign, brutal, civil, comical, congenial, connubial, conventional, cordial, criminal, equal, familiar, fatal, final, formal, frigid, frugal, general, genial, hospital, human, immortal, impartial, individual, inhuman, insipid, intellectual, intrepid, irregular, jocular, jovial, legal, liberal, literal, lucid, moral, mortal, original, partial, particular, peculiar, personal, placid, popular, potential, principal, public, punctual, rapid, regular, rigid, sentimental, similar, singular, social, solemn, special, spiritual, stolid, stupid, timid, tranquil, uniform, vital, vulgar.

В нашем движении вниз по ранжированному списку пар протонаффиксов мы наталкиваемся на пару -ies – -y, последний протонаффикс связан также с -ied, -ily и -y's, а протонаффикс -ies связан с -y's и -ied. В свою очередь -ied связан с -ying, а -ily – с -ier.

	y's	ies	y	ied	ying	ily	ier
y's		62	55	-	-	-	-
ies	62		207	47	34	-	-
y	55	207		48	26	62	20
ied	-	47	48		145	-	-
ying	-	34	26	145		-	-
ily	-	-	62	-	-		47
ier	-	-	20	-	-	47	

Возникает странная (неоднородная) парадигма, в которой хвосты аффиксов совпадают со многими аффиксами в предшествующих трех парадигмах. Остается предположить какое-то чередование i/y, относящееся к основе. Тогда аффиксами станут -'s, -es, -0, -ed, -ing, -ly, -er, которые и распределятся по трем агрегированным парадигмам:

1) adversary, apothecary, beauty, century, charity, cherry, chivery, company, country, destiny, family, functionary, history, hospitality, landlady, liberty, luminary, memory, misery, mulberry, nobody, property, secretary, security, sentry, society, soliloquy, somebody, vestry – 29 основ, сочетающихся с тремя аффиксами, + мощная пара аффиксов, появляющаяся у 170 основ: ability, ecstasy, missionary, novelty, rosary, trophy, etc.

2) accompany, exemplify, gratify, identify, intensify, justify, magnify, satisfy, terrify, testify (по 4 аффикса); beautify, certify, comply, diversify, electrify, fortify, mollify, mortify, multiply, mystify, occupy, prophesy (!), qualify, signify, specify, supply (по 3 аффикса).

3) crafty, dainty, gloomy, mighty, pretty, scanty, steady, thirsty, weighty, worthy (по 3 аффикса).

На чередование букв с/t указывают связи следующей матрицы:

	tial	ce	ced	ces	t	tly	cy
tial		50	-	24	18	34	-
ce	50		55	112	115	84	12
ced	-	55		67	8	-	-
ces	24	112	67		31	37	-
t	18	115	8	31		123	64
tly	34	84	-	37	123		89
cy	-	12	-	-	64	89	

По 5 аффиксов имеют основы confiden-, consequen-, distan-, excellen-, inconvenien-, reveren-; по 4 аффикса commen-, convenien-, conwenien-, defian-, differen-, elegen-, eloquen-, eviden-, extravagan-, ignoran-, impertinen-, incoheren-, independen-, influen-, insolen-, instan-, intelligen-, observan-, peniten-, permanen-, prominen-, providen-, pruden-, significan-, silen-, subservien-. Отсутствие нулевого аффикса в этом наборе (а значит – исключительно связанные основы) указывает на морфологию какого-то неанглийского типа.

Особую парадигму образуют 6 аффиксов, каждый из которых появляется и как отдельное графическое слово /-five, -four, -nine, -one, -six, -two/. Учитывая последнее обстоятельство, по-видимому, имеет смысл говорить здесь о словосложении. Такую же трактовку примем и для набора /-0, -man, -men/: эксклюзивной парадигмой здесь выступают основы china, clergy, cracks, dutch, english, french, hunts, irish, lands, lighter, marks, sports, statesman, swords, trades; по 3 аффикса имеют также barge, chair, coach, country, gentle, guards, highway, horse, husband, india, journey, noble, plough, police, towns, watch, water.

Отметим три словообразовательных аффикса, присоединяемых к субстантивам.

/-0, -ful, -s/

Эксклюзивную парадигму составляют churchyard, deceit, spade, spite, tea-spoon, teacup, thimble, thowt. Все три аффикса соединяются с основами: basin, basket, boast, bucket, care, change, cheer, delight, disdain, disgrace, distrust, doubt, dread, event, fear, forget, fret, fright, fruit, glee, grace, grate (!), hate, health, hope, kettle, mistrust, mourn, mouth, neglect, pain, plate, play, power, regard, regret, resent, respect, revenge, right, scoop, scorn, skill, sorrow, spoon, shame, taste, tear, teaspoon, thank, thought, trust, truth, tumbler, tune, waste, wonder, wrong, youth²⁵.

/-0, -less, -s /

Эксклюзивная парадигма представлена всего шестью основами bottom, defence, dollar, expression, lustre, portion. Еще какие-то аффиксы имеют 44 основы: blossom, boffin, breath, brother, button, carpet, change, colour, comfort, conversation, emotion, exhaust, father, fathom, feather, finger, friend, ground, handle, meaning, measure, mother, motion, number, object, passion, pasture, pattern, purpose, regard, relent, remorse, resist, shadow, shelter, speech, spirit, spring, stocking, supper, tenant, thought, whisker, window.

/-0, -'s, -ship, -s /

В эксклюзивную парадигму входит восемь основ: chancellor, emigrant, godmother, guardian, member, messenger, partner, proprietor. По крайней мере три аффикса соединяются с 20 основами: 'prentice, author, bachelor, beadle, companion, cousin, director, editor, fellow, friend, general, leader, lord, master, prison (!), relation, seaman, secretary, steward, workman.

²⁵ С учетом сказанного относительно i/y в этом списке оказались бы также beauti-, duti-, merci-, piti-, plenti-. Снятие порога на длину основы добавило бы art, fit, joy, law, man, use, woe. Большинство производных с -ful относится к адъективам, что доказывается на стадии уточнения парадигм, когда в игру вступает пятибуквенный аффикс -fully. Приходится ожидать перехода к минимальному интервалу, чтобы убедиться в субстантивном характере слов basinful, basketful, bucketful, churchyardful, kettleful, mouthful, plateful, spadeful, spoonful, teacupful, teaspoonful, thimbleful, tumblerful.

Как и в русских корпусах, на первом месте в списке пар префиксов стоит отрицательный префикс /0 – un, S=233/, соединяющийся как с адъективными, так и с глагольными основами (всего 732 основы)? ср. *accountable, artistic, blessed, ceremonious, civil, earthly, equal, faithful, feeling, generous, healthy, heroic, likely, lucky, natural, poetical, thankful, womanly, worldly*, а также *clasp, clench, cover, deceive, fasten, settle, winds*. Иногда этот префикс присоединяется к субстантивным основам, образуя глаголы, что, впрочем, будет доказано на минимальном интервале, ср. *earth, chain, robes, yoked*. На втором месте в ранжированном списке стоит пара /0 – dis, S=73/, в эксклюзивной парадигме 143 основы, ср. *advantage, agreeable, appoint, approve, believe, charge, content, credit, embark, figure, grace, honest, inherit, reputе, similar, taste*. Иногда *dis-* и *mis-* соединяются с одной и той же основой, например, с *trust, trustful*. Последний префикс сочетается с *apprehend, behaved, carried, conception, construction, deeds, directing, fortune, guided, interpret, pronounced, quotation, understand* и т. п.²⁶

Далее следует пара /0 – a-, S=66/, где префикс *a-* соединяется с глагольными основами с исходом на *-ing*: *beginning, blowing, buying, cooking, goin', kissing, nursing, shaking, sitting, trembling* и др. Орфографически *a-* варьирует с *a*, оба варианта сочетаются с основами *begging, dying, giving, going, larning, laughing, listening, lyin', musing, piece, waiting*.

Есть еще три префикса, создающие глаголы и значимо связанные только с 0-. Это *en-* (ср. *chanted, circle, countered, danger, forced, large, listed, slave, tangle, titled* и др.); *out-* (ср. *breaks, burst, casts, lived, rages, skirts, stretched, weigh* и т. п.); *be-* (ср. *calmed, decked, devilled, friend, grudge, labour, moaned, smeared, spake, speak, spoke, stirred, stowed, wilder* и т. д.)²⁷.

Остальные префиксы так или иначе входят в набор латино-романских моделей и потому дистрибутивно связаны друг с другом.

	0	com	con	de	dis	ex	im	in	per	pre	pro	re	sup
0		17	17	24	73	14	41	60	6	20	9	58	8
com	17		-	-	15	28	45	-	-	3	14	17	43
con	17	-		-	19	17	-	46	26	27	21	38	-
de	24	-	-		10	19	22	25	11	15	30	40	11
dis	73	15	19	10		27	19	7	-	10	10	36	10
ex	14	28	17	19	27		33	10	-	8	24	15	20
im	41	45	-	22	19	33		-	-	-	10	23	51
in	60	-	46	25	7	10	-		21	20	15	29	-
per	6	-	26	11	-	-	-	21		-	-	17	-
pre	20	3	27	15	10	8	-	20	-		6	24	-
pro	9	14	21	30	10	24	10	15	-	6		17	6
re	58	17	38	40	36	15	23	29	17	24	17		15
sup	8	43	-	11	10	20	51	-	-	-	6	15	

Примерами сочетаемости основ и аффиксов служат:

posing	com de dis ex im pro re sup
position	0 com de dis ex im pro sup
press	0 com de ex im re sup
sisted	con de in per re
ference	con de in pre re
verted	con in per re
tested	0 con de pro
duction	de in pro re
scribed	de in pre pro
formed	0 de in per re
covers	0 dis re
pound	0 com ex pro

²⁶ Любопытно, что все три префикса имеют семантическое сходство – в широком смысле – отрицание.

²⁷ Есть несколько случаев деэтимологизации *be-* в служебных словах – *because, betides, betimes, between, betwixt*.

В большинстве случаев собственная семантика этих префиксов затуманена, однако о сохранении такой семантики может свидетельствовать большое число основ при префиксах, соотнесенных с 0-. Примерами явятся такие пары:

0 de	48 основ:	block, cipher, generate, grade, merit, privation;
0 im	68 основ:	material, measurable, memorial, moderate, modest, moral, mortal, partial, patient, perfect, personal, piety, politic, probable, provident, prudent;
0 in	187 основ:	accessible, active, audible, capable, clemency, comparable, competent, complete, constant, convenient, correct, credulous, decision, definite, delicate, dependent, direct, discretion, distinct, divisible, effective, effectual effectually, famous, fidelity, formal, gratitude, hospitable, human, justice, secure, sensible, stability, tolerable, valid, voluntary;
0 re	137 основ:	building, commence, consider, kindle, modelled, paying, write. ²⁸

Вторым корпусом текстов для ДСА послужили драматические произведения Шекспира (F = 850 тыс. словоупотреблений). Поскольку этот корпус намного меньше, чем корпус Диккенса, мы можем ожидать меньших значений S. Действительно, начало ранжированного списка дает вдвое меньшие значения S.

215	0	s	88	man	men	82	d	st
143	'er	ver	87	d	s	82	ed	ing
111	's	0	86	'	0	75	0	ly
110	ies	y	84	r	st	73	's	s
90	'd	ed	83	d	r	73	0	d

В корпусе Шекспира преобладает поэзия, а потому для него характерны варианты с опущенным слогом, передаваемым на письме апострофом. Типична протопарадигма /'er – ver/ с основами howsoe-, moreo-, whate-, whatsoe-, whene-, wheresoe-, whosoe-.

Морфологическая система, статистически открываемая в текстах Шекспира, очень похожа на то, что было получено на корпусе Диккенса. Главное отличие видно в глагольной парадигме. Обратимся к матрице внутренних связей²⁹:

	0	'd	e	ed	es	est	eth	ing	s	'st	er	ion
0		58	–	54	14	14	13	55	(215)	25	22	13
'd	58		33	90	35	15	23	72	44	11	22	3
e	–	33		47	58	11	13	40	–	–	11	30
ed	54	90	47		43	17	23	82	42	17	25	23
es	14	35	58	43		14	18	36	–	–	11	13
est	14	15	11	17	14		10	21	8	20	(31)	–
eth	13	23	13	23	18	10		32	11	–	6	4
ing	55	72	40	82	36	21	32		46	26	31	10
s	(215)	44	–	42	–	8	11	46		19	16	5
'st	25	11	–	17	–	20	–	26	19		(17)	–
er	22	22	11	25	11	(31)	6	31	16	(17)		–
ion	13	3	30	23	13	–	4	10	5	–	–	

Здесь мы видим особые аффиксы –est (с вариантом –'st) и –eth, отсутствующие в глагольной парадигме Диккенса. Вот некоторые основы и сочетающиеся с ними аффиксы:

²⁸ Интересен контраст с русскими провербами. Старая (индоевропейская) система глагольных префиксов выветрилась в латино-романских заимствованиях. Но в английском языке число заимствованных основ настолько велико, что необходимость в создании антонимов привела к сохранению продуктивности im- и in- при таких основах. В какой-то степени сохраняет свою продуктивность re- с новой (и актуальной) семантикой повторяемости.

²⁹ Взяты в скобки омонимичные пары из адъективной парадигмы –er и –est, а также –0 – –s из субстантивной парадигмы.

speak	0 'st 't er est eth ing s
stand	0 'st ard (!) est eth ing s
esteem	0 'd 'st ed est eth s
perish	0 'd ed en est eth ing
bless	0 'd ed es eth ing
reach	0 'd es eth ing
bring	0 'st er est eth ing s
build	0 ed eth ing s
chang	'd e ed es est eth ing
deceiv	'd e ed er es est eth ing
resolv	'd e ed es eth.

Третьим английским корпусом для проверки ДСА послужил комплект газеты «The Times» за 1995 г. (F = 45 миллионов словоупотреблений) – своего рода аналог русского корпуса «Независимой газеты». Из соответствующего частотного словаря отобраны графические слова с частотой 10 и более (примерно 70 тысяч разных слов).

Верхушка ранжированного списка этого корпуса самым поразительным образом совпадает с соответствующим списка корпуса Диккенса. Исчезли только числительные (эффект слитного написания, нередкого в XIX в.) и пара /' – g/, почти невозможная в современной газете.

496	0	s	184	ies	y	131	es	ing
359	0	's	178	d	s	129	ce	tly
312	ed	ing	164	ng	ons	126	e0	ing
206	ied	ying	159	0	d	120	0	ing
202	d	r	152	on	ve	118	0	ed
199	d	rs	138	s	's	115	ed	ion
197	ng	on	138	ly	ness	115	ons	vely
190	0	ly	134	on	ve	106	ed	s

Три главные агрегированные парадигмы сохраняются в неприкосновенности. Как пример рассмотрим внутренние связи глагольной парадигмы:

	0	ed	ing	s	er	ion	ive	ment	or
0		118	120	(496)	41	20	11	20	8
ed	118		312	106	66	115	38	40	46
ing	120	312		105	84	94	34	31	47
s	(496)	106	105		34	14	11	26	8
er	41	66	84	34		-	-	-	-
ion	20	115	94	14	-		94	-	87
ive	11	38	34	11	-	94		-	51
ment	20	40	31	26	-	-	-		-
or	8	46	47	8	-	87	51	-	

Небольшие различия отмечены у некоторых словообразовательных аффиксов. В три раза сократилось число основ в парадигме /-0, -ful, -s/: deceit, disdain, faith, force, forget, fruit, grace, grate, insight, meaning, regret, resource, shame, sorrow, succes, truth, worship, youth. Особенно заметно сокращение числа основ – наименований сосудов: bucket, spoon, teaspoon. Сократился и набор основ с аффиксами /-0, -less, -s/: bottom, breath, character, colour, defence, direction, effort, emotion, expression, father, flight, friend, ground, leader, meaning, mother, motion, motive, number, passion, purpose, regard, relent, sleeve, thought, ticket, victim, weight, wicket, window.

С другой стороны, заметными стали слова с суффиксом -ee: absentee, addressee, appointee, assignee, deportee, employee, escapee, grandee, interviewee, moralee, referee, trainee, transferee, trustee (в корпусе Диккенса находим absentee, debauchee, divorcee, licensee mortgagee, nominee, trustee).

2.4. Французский язык

Полигоном для открытия французской морфологии послужили романы писателей XIX века: В. Гюго (3 романа), Г. Флобера (3), А. Дюма (4), А. Франса (3), Э. Сю (2), Г. Мопассана (2), А. Стендаля (1), Ж. Санд (1), Ж. Верна (1), братьев Гонкуров (1). Общий объем корпуса составляет 4,5 млн. словоупотреблений.

Снова ранжируем список пар протоаффиксов.

673	0	s	148	0	e	107	0	ment	91	moi	nous
293	-elle	-il	147	e	s	107	a	er	89	a	é
275	elle	il	143	f	ves	106	ce	tes	88	-le	-moi
273	ent	t	141	0	it	104	ant	er	87	ant	é
217	it	nt	139	0	es	101	er	ée	87	nt	z
191	f	ve	138	fs	ves	100	ce	ts	86	0	-vous
184	-il	-on	136	0	nt	100	moi	vous	86	ait	e
184	ient	nt	131	ce	te	99	0	ient	86	ait	ent
164	es	s	118	er	é	97	ce	t	86	ient	nte
149	cinq	huit	115	fs	ve	96	r	z	86	nt	r
149	cinq	sept	114	0	r	95	ez	ons	85	ant	ent
149	cinq	six	112	ait	er	94	er	és	83	ait	ez
149	huit	sept	111	f	on	93	ait	é	83	ient	ntes
149	huit	six	111	s	t	92	it	nte	82	ait	ée
(149)	sept	six	108	ment	s	91	eur	ion	81	as	ont

Как и в случае английского языка, первой оказывается пара 0 – s, соответствующая эксклюзивная протопарадигма объединяет 5476 основ. Далее следуют пары с протоаффиксами -elle, -il и -on, из 45 соответствующих основ 42 основы оканчиваются на -ait. Знак дефиса неизменно предшествует протоаффиксу в парах moi – vous, moi – nous, le – moi. Выстраивается парадигма с 8 аффиксами: -la, -le, -les, -lui, -moi, -nous, -vous, -y. Внутренние связи в этой парадигме представлены в следующей матрице.

	la	le	les	lui	moi	nous	vous	y
la		85	55		47		26	46
le	85		47	42	88	26	39	31
les	55	47			32			
lui		42			55		38	
moi	47	88	32	55		42	47	21
nous		26			42		33	
vous	26	39		38	47	33		
y	46	31			21			

По 7 аффиксов присоединяют к себе faites, laissez; 6 аффиксов у основы prenez; 5 аффиксов у основ appelez, mettez, prenez, regardez.

Ядро глагольной парадигмы формируется 9 аффиксами: -0, -ble, -ient, -is, -it, -nt, -nte, -r, -z.

	ble	ient	is	it	nte	nt	0	r	z
-ble		43	23	46	9	30	11		
-ient	43		(105)	(273)	86	184	99		
-is	23	(105)		(139)	21	58	45		
-it	46	(273)			92	217	141		
-nte	9	86		92		92	20		
-nt	30	184		217	92		136	86	87
-0	11	99		141	20	136		114	63
-r						86	114		96
-z						87	63	96	

Бросается в глаза внутренняя неоднородность этой парадигмы: четыре первых аффикса неизменно присоединяются к основам с исходом на а-, аффиксы -nt и -0 также могут сочетаться с такими же основами. Все семь аффиксов /-ble, -ient, -is, -it, -nte, -nt, -0/ сочетаются с основой méprisa. У 21 основы находим 6 аффиксов из этого набора (accepta, brûla, compta, considéra, déchira, désira,

détesta, mangea, passa, regretta, remarqua, respecta и др.). Некоторые мощные протопарадигмы включают 4 аффикса набора; примером может служить /-0, -ient, -it, -nt/, содержащая 213 основ (acheta, approuva, commanda, dressa, opposa, reporta и др.).

Два последних аффикса неизменно присоединяются к основам с исходом на е-, аффиксы -nt и -0 также могут сочетаться с ними. Эксклюзивная парадигма с четырьмя аффиксами включает 25 основ (aborde, admire, agite, exprime, imite, respire, souhaite и т. п.). Общее же число основ, сочетающихся со всеми 4 аффиксами, достигает 277 (chante, cherche, compte, couche, demande, désire, embrasse, écoute, frappe, garde, laisse, mange, monte, montre, pardonne, présente, regarde и т. п.).

Как видим, в двух частях парадигмы есть основы (compta/e, mange(a)), различающиеся лишь конечной буквой. Это заставляет подозревать, что в дальнейшем обнаружатся более длинные аффиксы. Действительно, уже в верхней части ранжированного списка встречаются пары ait/ent, ait/er, ait/ez, ant/ent, ant/er. Появляется второй глагольный кластер с 14 аффиксами (-a, -ai, -aient, -ais, -ait, -ant, -as, -ent, -er, -é, -ée, -ez, -ons, -ont).

	a	ai	aient	ais	ait	ant	as	ent	er	ez	é	ée	ons	ont	0
a		76	62	56	130	94	44	66	107	79	89	80	59	50	31
ai	76		19	83	62	19	77	20	16	71	16	17	66	78	32
aient	62	19		29	81	75	7	62	56	40	46	39	30	15	6
ais	56	83	29		60	36	45	36	26	72	22	23	63	49	19
ait	130	62	81	60		116	32	86	112	83	93	82	61	43	26
ant	94	19	75	36	116			85	104	64	87	78	45		4
as	44	77	7	45	32					48			39	81	15
ent	66	20	62	36	86	85			73	60	64	63	36	5	27
er	107	16	56	26	112	104		73		53	118	101	31		
ez	79	71	40	72	83	64	48	60	53		52	53	95	47	15
é	89	16	46	22	93	87		64	118	52		125	32		
ée	80	17	39	23	82	78		63	101	53	125		32		
ons	59	66	30	63	61	45	39	36	31	95	32	32		39	15
ont	50	78	15	49	43		81	5		47			39		19
0	31	32	6	19	26	4	15	27		15			15	19	

С этой парадигмой ассоциированы еще четыре аффикса -е (S=40), -era (S=40), -ât (S=37) и iez (S=39). Поучителен пример аффикса era, связанного с erai (S=50), erais (S=25), erait (S=59), eras (S=31), erez (S=46), erons (S=23), eront (S=22). Хвосты этих аффиксов выступают как доказанные аффиксы 14-членной парадигмы, отсюда следует вывод: суффикс -er является основообразующим суффиксом внутри глагольной парадигмы.

Обратим внимание на то, что -0 слабо связан с данным кластером³⁰.

Не обнаружено ни одной основы, которая сочеталась бы со всеми 14 аффиксами. Сочетаются с 12 аффиксами 7 основ (cherch, demand, demeure, racont, retour, retrouv, touch); по 11 аффиксов у 17 основ (compt, conserv, contempl, désespér, laiss, manqu, montr, pleur, pouss, quitt, remont, rentr, réveill, soupçonn, travaill, trembl, trouv); десять аффиксов видим у 21 основы (approch, couch, déclar, expliqu, fâch, prouv и т. п.), 9 аффиксов у 46 основ (admir, arriv, consol, examin, frapp, observ, tourn, visit и т. д.).

³⁰ Присутствие -0 при основах наряду с нашими 14 аффиксами объясняется реальным существованием субстантивов и адъективов, ср. présent (наряду с 12 положительными аффиксами), chant, continu, devin, rapport, recul, regard, souhait (11), arrêr, désir, mépris, oubli (10), appel, apport, calcul, écart, goût, prêt, profit, rappel, refus, repos, respect, support, tourment, trait (9). В этих случаях -0 – реальный нулевой аффикс.

По-другому обстоит дело с основами на -r (инфинитивами), ср. demander, finir, laisser, partir, sortir, trouver (9), gagner, garder, manger, manquer, marcher, monter, montrer, passer, quitter, regarder, retrouver, sentir, servir, tirer (8), aimer, changer, contenter, continuer, donner, parler, payer, rester, retourner, sauver, trahir, venger (7). То, что выше было сказано о суффиксе -er, приводит нас к другому выводу относительно -0, последний должен быть квалифицирован как псевдоаффикс.

Двигаясь вниз по ранжированному списку, обнаружим две узкие парадигмы: 1) /-f, -fs, -ve, -ves/ со средним значением S=142; все основы имеют исход на i-. В эксклюзивную протопарадигму входит 7 основ (chéti, crainti, excessi, fugiti, plainti, primiti, rétrospecti). По 4 аффикса имеют еще 13 основ, из которых 7 одновременно сочетаются с -on и -ons (attenti, convulsi, expressi, positi, relati, significati). 2) /-cinq, -huit, -sept, -six/ со средним значением S=149 и всего с четырьмя основами: quatre-vingt-, cinquante-, quarante-, soixante-.

Следующая узкая (но очень мощная – более 1000 основ) парадигма включает четыре аффикса (-0, -e, -es, -s). Эксклюзивная четырехчленная парадигма объединяет 805 основ, эксклюзивная парадигма с аффиксами /-0, -e, -es/ объединяет 226 основ. Эта парадигма может быть названа адъективной; характерные основы – absolu, allemand, anglais, aucun, civil, court, direct, divin, espagnol, exact, écossais, fatal, final, français, froid, futur, gaulois, grand, japonais, latin, lourd, mauvais, meilleur, méchant, petit, plein, profond, saint, sourd, vacant. Однако и субстантивные основы здесь тоже представлены, ср. courtisan, cousin, libertin, marquis, martyr, sultan, voisin. Среди основ численно преобладают отглагольные адъективы («причастия») – как доказанные регулярные с исходом на é-, ble- и ant-, так и более редкие с исходом на u- (attendu, confondu, descendu, devenu, défendu, entendu, perdu, prétendu, suspendu, vendu и др.), -is (admis, compris, entrepris, promis, remis, requis, surpris, etc.) и it- (conduit, construit, décrit, défait, détruit, interdit, maudit, prescrit, produit, satisfait, séduit, etc.). Две названные протопарадигмы дают возможность выделить часто повторяющиеся хвосты основ – кандидаты в словообразовательные суффиксы:

ent- (ardent, différent, éloquent, éminent, évident, fervent, imprudent, indigent, indulgent, innocent, insolent, intelligent, récent, transparent, turbulent, urgent, véhément, etc.);

ain- (contemporain, hautain, humain, incertain, lointain, malsain, mondain, napolitain, puritain, romain, surhumain, vilain).

Особого внимания заслуживает исход на al-: наряду с ожидаемыми аффиксами у основ banal, fatal, final, здесь появляется парадигма /-al, -ale, -ales, -aux/ у основ anim, besti, capit, cordi, féod, génér, horizon, immor, impart, impéri, inéga, inferna, latéra, libér, maréch, méridion, municip, nation, occident, orient, origin, princip, provinci, sacerdot, septentrion, sign, soci, spéci, transvers, vertic, végét. Возможность чередования ux/l позволяет отнести к тому же классу и парадигму /-use, -uses, -ux/. Все три аффикса сопровождают 31 основу, неизменно оканчивающуюся на e-: amoure, aventure, calomnie, courage, ennuye, fâche, harmonie, hasarde, nombre, paresse, savoure, scandale, série и т. п.

К мощной адъективной парадигме примыкает парадигма /-0, -ment, -s/, в эксклюзивной парадигме обнаруживаем 318 основ, на 98% оканчивающихся на e- (active, amoureuse, aucune, certaine, chaude, civile, cordiale, courageuse, divine, fatale, finale, générale, génèreuse, harmonieuse, héroïque, heureuse, libérale, magique, militaire, morale, naturelle, parfaite, personnelle, pleine, publique, relative, royale, sérieuse, sentimentale, sourde, tacite, unique, verticale, vulgaire, etc). Еще 50 основ могут присоединять и другие аффиксы (absolu, aveuglé, chaste, double, drôle, grande, infini, ingénu, longue, misérable, résolu, etc). У парадигм /-0, -e, -es, -s/, /-f, -fs, -ve, -ves/, /-use, -uses, -ux/ и /-0, -ment, -s/ находим множество сходных основ, незначительно различающихся своими хвостами, откуда следует лишь один вывод: все это одна большая адъективная парадигма³¹.

В верхней части ранжированного списка остались три пары с - t, с - te, с - tes, составляющие ядро парадигмы /-ce, -ces, -t, -te, -tes, -ts/ со средним значением S=98. Соответствующие основы кончаются на n-. Полный набор шести аффиксов представлен у 25 основ (absen, apparen, complaisan, différen, excellen, élégan, ignoran, impatient, innocen, intelligen, péniten, puissan, violen и др.),

³¹ По-видимому, другой парадигмой следует считать набор /-0, -ment, -ments, -s/ с основами attache, balance, commande, commence, contente, embrasse, enseigne, escarpe, ménage, renonce, ressenti, tremble. Этот вывод, формулируемый как гипотеза на этапе микроинтервала, превратится в полную уверенность на минимальном интервале.

5 аффиксов – у 46 основ (abondan, ambulan, confian, confiden, corresponдан, croyan, distan, éloquen, fréquen, importan, influen, instan, naissan, nonchalan, patien, plaisan, présen, pruden, vacan и пр.), 4 аффикса у 18 основ (clairvoyan, concurren, conséquen, descendan, enfan, méfian и др.). К шестичленной парадигме примыкает семичленная парадигма /-mmen, -nce, -nces, -nt, -nte, -ntes, -nts) с 19 основами (complaissa, conséque, diffère, excelle, éléga, éloque, indépenda, intellige, nonchala, suffisa и др.).

На грани верхушки ранжированного списка находим пару eur – ion – ядро субстантивной парадигмы /-eur, -eurs, -ion, -ions, -rice/, где все пять аффиксов сочетаются с основами direct, indicat, institut, persécut, protect. По 4 аффикса имеют более 30 основ (accusat, collect, composit, confess, connaiss, conspirat, créat, élect, invent, libérat, observat, opérat, profess, protect, provocat, success, traduct и др.).

В ранжированном списке пар протоаффиксов остается еще 12 пар, чье S превышает 50. За единственным исключением (au – aux) все они имеют вид (X – X+s). Поскольку дистрибутивное сходство s и x было замечено на примере основ с исходом на -al, эти 13 пар можно считать вариантом самой мощной («именной») пары /-0, -s/, возглавляющей ранжированный список:

- age – 23 основы (bailli, d'ouvr, espionn, échafaud, équip, feuill, hérit, marchand, nettoiy, personn, témoign и др.);
- aire – 18 основ (aliment, capitul, célibat, comment, commiss, doctrin, élément, fonctionn, humanit, incendi, mousquet, nécess, parlement, propriét, témér и др.);
- ard – 7 основ (boulev, brouill, campagn, étend, montagn, traîn, vieill);
- au – 26 основ (bande, barre, berce, boule, bourre, drape, flambe, gâte, mante, moine, panne, perdre, pince, roule, table, tombe, troupe и т. п.);
- et – 22 основы (banqu, bouqu, chapel, cheval, décr, feuill, fleur, gourm, oeill, pistol, poign, préf, robin souffle и т.п.);
- ien – 7 основ (autrich, égypt, grammair, paroiss, pratic и т.п.);
- ier – 43 основы (aventur, banquerout, bijout, boutiqu, cabaret, canonn, carabin, charpent, cheval, contreband, cordonn, cuirass, cuisin, dragonn, financ, gondol, grenad, hérit, jardin, journal, manufactur, meurtr, particul, perruqu, serrur, tapiss, tonnel и т.п.);
- iste – 6 основ (bonapart, buonapart, capital, cardinal, journal, légitim);
- ité – 18 основ (célébr, complic, importun, inégal, libéral, médiocr, nécess, original, prospér, univers и т.п.);
- erie – 22 основы (artille, bouche, coquette, espiègle, flatte, galante, gauche, jacque, ménage, parfume, plaisante, raille, sauvage, trompe и т.п.);
- tude – 3 основы (inquiét, servit, sollicit)
- ure – 35 основ (architect, chauss, confit, couvert, créat, embras, écrit, fournit, garnit, nourrit, ouvert, procéd, sculpt, signat и т.д.

Ранжированный список пар протопрефиксов очень невелик.

600	d'	l'	238	m'	s'	143	0	dé	112	d'	m'
511	0	l'	158	0	m'	140	d'	s'	108	l'	t'
429	0	d'	167	0	re	133	n'	s'	105	j'	l'
288	l'	s'	165	l'	n'	132	0	n'	104	s'	t'
269	0	s'	158	0	m'	121	j'	n'	100	ble	gli
251	l'	m'	158	m'	t'	116	has	reg			

Некоторые из протопарадигм объединяют десятки и сотни основ (0 l' – 791, 0 d' – 531, 0 s' – 281, 0 m' – 82, 0 n' – 85). Список возглавляют (0, d', j', l', m', n', s', t'). Графическое своеобразие второго знака этих протопрефиксов (') и то обстоятельство, что им точно соответствуют чрезвычайно частые слова с конечным -e [de (f=197438), je (f=42452), le (f=105567), me (f=14671), ne

(f=35148), se (f=32124), te (f=3113)], заставляет видеть в этих префиксах проклитики, обусловленные началом основы (a-, e-, é, ê-, h-, i-, o-, ô, u-). Максимальный набор положительных префиксов (6) видим у основ: abandonne, accorde, adore, aimer, appelle, approche, attendre, emporte, entende, entendre, éprouvante, être, occupe, ouvre, ouvrir. По пять префиксов обнаружено у 60 основ (accuse, adresse, aller, arrête, avoir, embrasse, estime, écouter, écrire, épouse, étais, instruire, offrir, oublier, ôter и др.).

Восемь протопрефиксов не образуют единой системы, но остаются отдельными словообразовательными аффиксами, присоединяемыми к основе:

- con- — 112 основ (centre, citoyens, damné, fondu, forment, frère, grès, jurait, naissances, sciences, sentez, vaincre и т. п.);
- dé- — 832 основы (bouchait, cacheter, chargent, coiffer, composer, couper, favorable, finir и т. п.);
- en- — 122 основы (chaîne, courage, chanté, durci, fouir, traîner, voler и т. п.);
- é- — 335 основ (changer, clair, criez, loigné, mouvoir, perdu и т. п.);
- im- — 114 основ (matérielle, mobile, monde, mortel, parfait, patient, payable, piété, portant, prévu, prudent, puissant и т. п.);
- in- — 377 основ (accoutumé, attendu, capable, certitude, compatible, croyable, décente, discret, fidèle, fortuné, habitable, juste, occupé, sensé и т. п.);
- re- — 1360 основ (bondir, boutonna, chercher, commencer, construit, devenir, douter, fouler, gagner, garder, lirai, mercier, montez, mords, parler, partie, peupler, tarder, tournent, trouver, vendeur и т. п.);
- sur- — 110 основ (chargée, excité, faces, humaines, lendemain, naturel, vivant и т. п.).³²

2.5. Немецкий язык

Проверка алгоритма осуществлялась на корпусе сочинений Э. Т. А. Гофмана (1160 тыс. словоупотреблений). Ранжированный список пар протопрефиксов выглядит следующим образом:

310	0	n	146	en	er	112	de	der	89	en	t
254	n	r	140	de	den	110	em	es	86	dem	der
221	e	en	136	er	es	108	en	es	83	dem	des
197	0	e	133	m	s	106	den	der	81	ste	sten
194	0	s	131	e	er	104	de	des	80	en	te
188	0	en	130	e	es	103	t	ten	78	n	nd
180	d	den	129	d	de	101	em	er	75	0	m
179	m	r	127	m	n	97	0	er	75	e	em
172	r	s	121	d	der	93	0	des	75	ene	enen
152	n	s	119	0	d	92	0	es	74	d	dem
147	0	r	117	t	te	91	den	des	73	em	en

Бросается в глаза обилие протопрефиксов с буквой e в предпоследней позиции графического слова, что создает дополнительную неопределенность при попытке интерпретации протопрефиксов как вариантов искомым аффиксов. В этих условиях особое внимание должно быть уделено хвостам основ. Начнем с шести однобуквенных протопрефиксов (-0, -n, -r, -s, -m, -e), связи между которыми отражены в следующей матрице³³.

	-0	-n	-r	-s	-m	-e
-0		310	147	197	75	197

³² Полностью ошибочными (случайными!) оказались пары (has- — reg-) и (ble- — gli), образовавшиеся благодаря реальным основам hasard/regard и blesser/glisser. Борьба с такими ошибками очень трудно. Остается утешаться тем, что доля подобных ошибок оказалась менее 1%.

³³ Пара 0/d присоединяется только к основам с исходом на n-, поэтому -d следует считать псевдоаффиксом. (См. -end в глагольной парадигме).

-n	310		254	152	127	-
-r	147	254		172	179	-
-s	197	152	172		133	18
-m	75	127	179	133		4

Первые пять аффиксов образуют хороший кластер. Соответствующая эксклюзивная («адъективная») парадигма объединяет 123 основы, оканчивающиеся на е- (andere, besondere, bessere, deine, dumme, eigne, ganze, goldne, große, harte, himmlische, kalte, kluge, kurze, lange, reine, seine, viele, wahre, zarte, etc.). У 31 основы помимо этих пяти аффиксов обнаруживаются еще какие-то аффиксы (bunte, ferne, feste, gleiche, graue, holde, keine, kleine, laute, leichte, leise, manche, milde, schlechte, schöne, starre, süße, tiefe, treue, volle, weite, wilde, etc.).

То обстоятельство, что все эти основы оканчиваются на е-, позволяет считать подлинными аффиксами -е, -en, -er, -es, -em. Такая интерпретация хорошо согласуется с нашим ранжированным списком, где представлены все десять ожидаемых пар. Но в том же списке видим пары 0/en, 0/er, 0/es, а за пределами списка также 0/em (S=52). Это кажется убедительным аргументом в пользу более широкой парадигмы /-0, -е, -en, -er, -es, -em/. Если обнаружатся основы с таким набором аффиксов, наша гипотеза превратится в уверенность. Действительно, соответствующая парадигма объединяет 49 основ (blank, blutig, dumpf, einiger, ernst, fremd, frisch, griechisch, himmlisch, hübsch, kimisch, physisch, plötzlich, ruhig, schwarz, solch, stark, teuflisch, trostlos, welch, etc.). Еще 36 основ к этому набору добавляют какие-то другие аффиксы (ähnlich, falsch, feierlich, freundlich, gering, häßlich, heillos, letzt, spitz, starr, wahrhaft, wirklich etc.). У 171 основы находим по пять аффиксов из шестичленного набора (deutsch, dünn, englisch, fatal, französisch, geistig, gnädig, heilig, mystisch, poetisch, scharf, schlimm, spät, spanisch, still, wichtig, etc.)³⁴.

К шестичленной адъективной примыкают еще 3 аффикса: -keit (среднее S=36), -ste (S=32), -sten (S=49). Полный набор девяти аффиксов имеют основы heftig, herrlich, lebendig, lieblich, seltsam, zierlich.

В ранжированном списке остаются теперь четыре пары протоаффиксов (t/te, t/ten, t/en, te/en), образующие ядро глагольной парадигмы, к которым еще добавляются пять аффиксов с меньшими значениями S.

	-t	-te	-ten	-en	-end	-ung	-n	-e	-st
-t		117	79	89	49	48	48	47	35
-te	117		103	80	49	50	40	35	14
-ten	79	103		44	36	31	20	26	15
-en	89	80	44		49	47	-	(221)	16
-end	49	49	36	49		26	-	35	15
-ung	48	50	31	47	26		5	25	5
-n	48	40	20	-	-	5		-	28
-e	47	35	26	(221)	35	25	-		17
-st	35	14	15	16	15	5	28	17	

При такой широте парадигмы трудно ожидать полного набора аффиксов у какой-либо основы. Восемь аффиксов обнаружено у двух основ äußer и erzähl. По 7 аффиксов присоединяют 13 основ (anschau, darstell, erblick, erinner, erreg, gewahr, rühr, schweb, spiel, stärk, stell, verfolg, versuch), по 6 аффиксов имеют 40 основ (drück, entfernen, erwach, fühl, glaub, kämpf, pfleg, schau, spring, streb, trenn, verdien, wiederhol, etc.); по 5 аффиксов находим у 98 основ (bedenk, beginn, besuch, bleib, bring, erkenn, erlaub, erschein, führ, könn, leich, schick, treib etc.).

Алгоритм не смог обнаружить субстантивные парадигмы. Из подозрительных узких протопарадигм упомянем /-0, -en/ с 464 основами, в т. ч. adjutant, advokat, akzent, aufstieg, bestand, chronik, figur, fragment, kamerad, musikan, nachtigall, notiz, person, psycholog, rezensent, rumor, sekundant, soldat, student, tendenz, violett, vortrag; albernheit, bosheit, gewohnheit, schönheit; kleinigkeit, schwierigkeit, süßigkeit; äußerung, andeutung, anmerkung, dictung,

³⁴ Содержательный анализ заставил бы отнести к субстантивам три основы — kleid, schmuck, tisch.

entdeckung, erzählung, meinung, prüfung, rechnung, richtung, sitzung, stellung, wohnung, zeitung (всего 47 основ с исходом на ung-); eigenschaft, herrschaft, narrheit, tollheit, wirtschaf, zarthheit; kabbalist, komponist, maschinist, organist, tenorist; dekoration, instruktion, komposition, konstellation, million, modulation, mystifikation, operation, reflexion, region, rezension, situation, variation, vision. Зная немецкий язык, исследователь признал бы субстантивами более 150 основ. Полностью субстантивной оказалась бы парадигма /-0, -s/ с 420 основами (bankier, dürer, frühling, garten, gluck, goethe, gouverneur, herzen, kardinal, karneval, lehrer, pater, shakespeare, tiger, vogel, etc.). Была бы признана субстантивной и парадигма /-0, -e (-en), -s / с 52 основами (abend, abgrund, anfang, angriff, ausruf, anschlag, augenblick, automat, charakter, gefecht, geisterreich, gericht, gespräch, getränk, instrument, jüngling, kampf, kostüm, leichnam, licht, lustspiel, manuskript, markt, meisterwerk, mittag, moment, mönch, offizier, organ, original, schicksal, schmuck, schrank, sonett, sonntag, staat, stein, stoff, strom, stück, sturm, talent, thron, trauerspiel, traum, umgang, unhold, unmut, verrat, vormittag, wacht, zweck).

Довольно легко выделяются немецкие префиксы. Относительно высокие значения обнаружены у шести префиксов, укладывающихся в единую парадигму

	0	be	ge	er	ver	ent
0		83	99	69	88	32
be	83		68	38	49	31
ge	99	68		44	77	29
er	69	38	44		32	15
ver	88	49	77	32		34
ent	32	31	29	15	34	

Все шесть префиксов соединяются с пятью основами (gangen, schlossen, setzt, standen, zogen). Комбинацию пяти префиксов (0-, be-, ge-, er-, ver-) встречаем у семи основ (folgt, kannt, richtet, schaffen, sehen, tragen, wiesen). Пять префиксов (0-, be-, ge-, ent-, ver-) также сочетаются с семью основами: (fallen, kleidet, nommen, schieden, setzen, stehe). Мощными оказались узкие протопарадигмы /0-, er/ – 112 основ (blickt, forsch, freuen, glüht, scheint, tönt, wänt, zählt, zeugt, etc.); /0-, be-/ – 131 основа (antwortete, darfst, durfte, fragte, fürchten, kämpfend, merkst, rechnung, rührten, sitzt, sorgte, stehend, suchst, treffend, etc.), /0-, ge-/ – 200 основ (bracht, dachten, fällt, hörten, kocht, lehrte, lobte, opfert, schickt, spielt, etc.). Громадное большинство основ, сочетающихся с перечисленными префиксами, относится к глаголам (ср. характерные суффиксы – -t, -ten, --st).

Вторая группа префиксов (числом 8) образуют не столь компактный кластер.

	0	ab	an	auf	aus	ein	vor	zu
0		32	53	42	34	40	36	24
ab	32		28	28	26	24	15	23
an	53	28		34	29	36	24	26
auf	42	28	34		31	28	21	22
aus	34	26	29	31		35	12	20
ein	40	24	36	28	35		18	20
vor	36	15	24	21	12	18		20

Все 8 префиксов сочетаются с основой nehmen; по 7 префиксов имеют основы geführt, gegangen, gehen, genommen, geschlagen, schlagen, sicht; по 6 префиксов – gerichtet, gestellt, gezogen, setzte; по 5 префиксов – brechen, einander, fälle, gefallen, gegeben, gelegt, gesetzt, getan, getragen, halten, hängen, nimmt, zuführen, zunehmen, zustellen, wenden.

Протопарадигмы с 0- обладают достаточной мощностью:

0- ab (37 gegessent, gestorben, sonderliche),

0- an (64 blick, erkannten, gedeutet, gehörte, gesicht, schaue, schlage, vertrauen),

0- auf (43 bewahren, enthalt, gefressen, geschlagenen, schreien)

0- aus (31 bleibe, geblieben, gelassener, gesprochen, sprache),

0- ein (43 beschert, dringt, geschlafen, hörner, studieren, verstanden),

0- vor (47 ausgesehen, bereitet, gestern, gesungen, trefflich)
 0- zu- (19 bereitete, gedacht, schauen).

Характерным образом префиксы второй группы («отделяемые префиксы») предшествуют префиксам первой группы.

Совешенно особняком стоит протопарадигма /0-, un-/ S=103 (178 основ angenehm, artig, bedeutend, bedingt, befangen, behaglich, bekannt, erwartet, geschick, gewiß, gewöhnlich, glück, heimlich, natürlich, nütz, ruhig, schuld, sinnig, vermutet, etc.).

Вторым объектом проверки алгоритма открытия парадигм стала газета «Süddeutsche Zeitung» за 1999 год (около 9 млн. словоупотреблений). При восьмикратном увеличении корпуса текстов, естественно, выросли и значения S пар протоаффиксов (от двух до четырех раз). Верхушка ранжированного списка дается ниже.

644	0	n	247	n	s	188	d	des	147	e	em
514	de	den	244	er	es	186	ene	enen	146	te	ung
514	n	r	240	den	der	180	m	s	145	ete	eten
506	e	en	239	sch	sche	174	em	er	141	um	ums
435	0	en	229	en	es	174	v	ver	140	ive	iven
400	0	e	229	t	ten	172	en	t	140	v	ves
399	0	s	223	em	es	170	ve	ver	139	ig	ige
303	d	de	221	r	s0	169	t	ung	139	iv	iven
298	de	der	219	0	er	168	ven	ver	138	en	te
271	e	er	213	ve	ven	168	ver	ves	138	en	ung
269	m	r	209	der	des	152	et	ete	138	ven	ves
264	0	r	207	den	des	152	ige	igen	136	igen	iger
263	t	te	205	m	n	152	nde	nden	135	et	eten
259	d	den	201	0	es	150	ium	iums	135	iv	ive
259	e	es	200	v	ven	150	rem	res	134	r	rn
259	en	er	200	tz	tzes	149	eibt	ieb	133	em	en
259	te	ten	194	d	der	148	re	ren			
253	de	des	192	v	ve	148	ve	ves			

Список на основе газеты мало отличается от списка на основе текстов Гофмана. К глагольной парадигме примыкает пара -et/-ete, в основном появляющаяся после t-, d- (abbild-, antwort-, anwend-, arbeit-, ausbild-, bedeut-, begegn-, belast-, bereit-, bericht-, betracht-, bezeichn-, breit-, dicht-, empfind-, ereign-, erwart-, gebet-, gericht-, gründ-, kleid-, leucht-, licht-, pflicht-, rechn-, schad-, schalt-, schneid-, schuld-, schüt-, spalt-, start-, stift-, verbind-, vermeid-, verwend-, verwund-, zünd-, etc.).

Увеличилось число пар с псевдоаффиксами: не только -d и -nd становятся частью -end, -v и -ve трансформируются в -iv (express-, intens-, mass-); -sch превращается в -isch (daemon-, histor-, islam-, olymp-, philosoph-, photograph-, serb-, soziolog-, telephon-).

Если учесть и другие пары с S>100, обнаружим подобное же расширение аффикса в парах: -on/-v > -ion/-iv (definit-, depress-, explosi-, informat-, pass-, posit-, etc., всего 56 основ); -k/-sch > -ik/-isch (analyt-, diplomat-, dramat-, elektr-, erot-, graph-, journalist-, kathol-, klass-, kulinar-, method-, paedagog-, plast-, polit-, takt-, tourist-, всего 65 основ). Появилась новая («субстантивная») парадигма /-erin, -ers, -ern, -er/ с основами angreif-, anhalt-, ansag-, arbeit-, autofahr-, berat-, berlin-, bild-, botschaft-, bürg-, dicht-, erzähl-, kämpf-, kassier-, lehr-, leit-, manag-, meist-, nachfolg-, politik-, reit-, richt-, schneid-, schreib-, schwimm-, sieg-, spiel-, sprech-, teilnehm-, unternehm-, vertret-, werf-, wien-, etc., всего 152 основы³⁵. Вариантом ее можно считать протопарадигму 0/in (S=68) с основами ausländ-, direktor-, kaiser-, künstler-, lektor-, maler-, minister-, nachbar-, priester-, schüler-, spion-, tschech-, türk-, zauber-.

³⁵ В корпусе Гофмана подобная парадигма включала бы 12 основ: общих с газетным корпусом – begleit-, besitz-, gesellschaft-, pfleg-, schauspiel-, spanirn-, tänz- и специфических гофмановских – dien-, rett-, säng-, träum-, verbrech-.

Яркая демонстрация торжества равенства полов в языке современной газеты.

Расширение статистически значимых пар протоаффиксов впервые вводит в сферу анализа чередования букв в основах, ср. -eibt/-ieb (S=149: liegenbl-, übrigbl-, übertr-, umshr-, untersch-, verschr-, zuschr-); -geht/-ging (S=112: einher-, hervor-, hinaus-, verloren-, voraus-, weiter-, zurück-); -abel/-able (S=111: akzept-, komfort-, praktik-, profit-, respekt-); -ürme/urm (S=111: kircht-, leuchtt-, wacht-, wirbelst-); -ieht/-og (S=108: durchz-, einbez-, überz-, unterz-, zurückz-).

Само собой разумеется, существенно различается лексическое наполнение парадигм. Например, протопарадигма /-0, -s, см. выше/ объединяет 1851 основу (accessoire, adenauer, airport, ankara, antonioni, apple, arafat, assad, atomtest, automobilclub, etc.). Парадигма /-0, -e (-en), -s/ включает теперь 174 основы (attentat, begriff, bezirk, dekret, dialog, exporteur, hormon, institut, kernkraftwerk, kinofilm, konzept, landtag, modell, objekt, parlament, point, projekt, skandal, staat, talent, etc.).

Система префиксов не различается существенно в двух корпусах. В газетном корпусе несколько усилились связи префикса um-. Появилась узкая парадигма /ex-, im-/ (S=63: matrikulation, portierte, ponierte, pressionsismus, pressionsisten).

2.6. Шведский язык

Тестируемым корпусом для шведского языка послужили тексты газеты «Göteborg Posten» за 1999 г. (3,15 млн словоупотреблений).

440	n	r	164	a	t	136	ig	iga	123	d	t
432	0	s	164	n	ns	135	en	er	121	d	ts
358	e	es	163	0	a	134	ad	ade	121	des	r
320	s	ts	162	sk	skt	134	et	ets	121	erar	erat
296	na	nas	157	an	r	134	va	vt	120	dd	dda
272	0	na	156	0	n	133	de	s	120	el	ella
262	0	en	156	ar	arna	133	re	ste	120	era	erar
232	de	r	150	and	öm	133	v	vt	120	erar	eras
200	de	des	150	gar	gas	132	ade	at	120	gde	ger
195	n	rna	149	d	des	130	0	et	118	de	ts
190	lsen	lser	148	en	ens	130	gad	gade	118	iga	igt
182	0	t	145	de	nde	130	isk	iska	118	rar	rat
180	erg	org	145	m	men	130	iska	iskt	117	n	t
180	or	orna	143	r	t	130	liga	ligt	116	ar	en
178	er	erna	142	ade	ar	130	rade	rat	116	v	va
175	de	t	142	ska	skt	126	des	s	115	ra	rar
170	0	e	141	st	ste	126	ig	igt	114	r	s
170	d	de	141	t	ts	125	0	r	113	ra	ras
168	des	ts	140	an	orna	125	en	erna	112	ra	rat
165	sk	ska	140	m	ter	124	0	er			

Матрица связей однобуквенных протоаффиксов создает довольно рыхлый кластер³⁶.

	s	n	r	0	d	t	a	e
s		-	114	432	100	95	-	-
n	7		440	156	-	117	-	-
r	19	315		125	99	143	-	-
0	3292	787	136		54	182	163	170
d	8	1	1	13		123	-	-
t	11	285	41	262	26		164	-
a	8	1	-	536	1	119		109
e	-	-	-	(495)	1	-	46	

³⁶ В отличие от симметричных матриц, рассмотренных в п. 3.2-3.5, в левой нижней части матрицы показаны не значения S, а число основ в эксклюзивных протопарадигмах.

В протопарадигме /-0, -e/ 322 основы оканчиваются на -ad, что позволяет заподозрить существование особого аффикса -ade (или -de). Пары протоеффиксов в данной матрице повторяются в составе пар большей протяженности: не только -0/-a, -a/-t, но и -ig/-iga, -iga/-igt, -sk/-ska, -ska/-skt, -v/va, -va/-vt. Так начинается вырисовываться ядро («адъективной») парадигмы с характерными аффиксами /-0, -a, -t/; как эксклюзивная парадигма она объединяет 429 основ (afrikansk, bisarr, blågul 'сине-желтый', borgerlik, burlesk, chanslös, emotionell, fiktiv, fruktbar 'плодотворный', genuin, global, judisk, kriminell, negativ, nervös, reell, riklig 'богатый', smutsig 'грязный', språklig 'языковой', sublim, svartvit 'черно-белый', tänkbar 'мыслимый', tandlös 'беззубый', temporär, total, vulgär, etc.). У 104 основ к этому трехчленному набору добавляется аффикс -e: aktuell, amerikansk, anonym, dansk, finländsk, finsk, fransk, grön 'зеленый', italiensk, liberal, lokal, massiv, militär, modern, norsk, österrikisk, politisk, polsk, spansk, suverän, värm, etc.³⁷

В матрице однобуквенных аффиксов особое положение занимает -d. Этот аффикс показывает сильные связи также с некоторыми дву- и трехбуквенными аффиксами; в результате выстраивается еще одна парадигма с особыми внутренними связями.

	d	de	des	nde	0	r	s	t	ts	re
d		170	149	66	54	99	100	123	121	27
de	170		200	145	89	232	133	175	118	52
des	149	200		80	60	121	126	111	168	32
nde	66	145	80		85	138	83	71	53	78
0	54	89	60	85		125	(432)	(182)	54	68
r	99	232	121	138	125		114	142	94	160
s	100	133	126	83	(432)	114		95	320	210
t	123	175	111	71	(182)	142	95		141	283
ts	121	118	168	53	54	94	320	141		44
re	27	52	32	78	68	160	210	283	44	

Все 10 аффиксов этой («глагольной») парадигмы встретились у 11 основ: avsluta 'закончить', berätta 'рассказать', betrakta 'рассмотреть', filma, måla 'красить', mörda 'убить', rädda 'спасти', skapa 'творить', spela 'играть', samla 'собирать', öppna 'открыть'. По 9 аффиксов имеют 45 основ (accipera, anklaga 'обвинить', arrangera, begränsa 'ограничить', beräkna 'вычислить', besluta 'решить', beteckna 'обозначить', bilda 'создать', döda 'убить', dominera, engagera, fascinerar, förmedla 'посредничать', granska 'проверять', grunda 'основать', handla 'торговать', hoppa 'прыгать', imponera, inspirera, kalla 'звать', komplettera, kontrollera, locka 'манить', minska 'уменьшать', operera, placera, planera, pressa, pröva 'пробовать', registrera, splittra 'расколоть', tillverka 'изготовить', träffa 'встретить', tvinga 'заставить', upprätta 'составить', uppskatta 'подсчитать', varna 'предостеречь', värdera 'ценить', etc.); по 8 аффиксов у 40 основ (arbета, avskaffa 'устранить', dansa, fånga 'поймать', förnya 'обновить', intervju, klara 'справиться', smitta 'заразить', spara 'копить', sparka 'ударить ногой', tappa 'терять', teckna 'чертить', testa, underteckna 'подписать', vårda 'заботиться', vänta 'ждать', älska 'любить', etc.).

Открывается и несколько субстантивных парадигм:

	0	ar	arna	en	ens
0		98	66	(262)	69
ar	98		156	116	39
arna	66	156		109	49
en	262	115	109		148
ens	69	39	49	148	

27 основ с полным набором аффиксов (eftermiddag, fredag 'пятница', kväll 'вечер', lördag 'суббота', sjukdom 'болезнь', tidning 'газета', tolkning

³⁷ Поскольку шведская лексика менее известна у нас, чем лексика английского, французского и немецкого языков, в данном параграфе многие основы сопровождаются переводом.

'перевод', ungdom 'молодежь', värld 'мир', etc.); 22 основы без -arna (drottning 'королева', häst 'лошадь', prins, torsdag 'четверг'); 108 основ без -ens; протопарадигма /-0, -en/ охватывает 1098 основ.

	0	er	erna	en	ens
0		124	87	(262)	69
er	124		178	135	46
erna	87	178		125	52
en	(262)	135	125		(148)
ens	69	46	52		

45 основ с полным набором аффиксов (artist, diktatur, fotograf, kampanj, kritik, provins, teknik). Вариантом этой парадигмы можно признать формы на -else (9 основ, например, händ 'случай', skrive 'деловое письмо'). Если бы не принятое ограничение на длину аффикса, к этим трем парадигмам присоединилась бы еще одна с 14 основами (arbetar, domar 'судья', fiskar 'рыбак', forskar 'ученый', läkar 'ученый', spelar 'игрок', etc.)³⁸.

	a	an	ans	or	orna
a		53	21	71	42
an	53		80	157	140
ans	21	80		57	-
or	71	157	57		180
orna	42	140	-	180	

Примерами служат fråg 'вопрос', människ 'человек', känsl 'чувство' (с пятью аффиксами), также flick 'девушка', kvinn 'женщина'.

Особняком располагается субстантивная парадигма («среднего рода»)

	0	et	ets
0		130	38
et	130		134
ets	38	134	

с 39 основами (institut, kokain, styck, etc.).

Последней парадигмой, заслуживающей внимания, является тройка /-re, -st, -ste/ («компаратив» и «суперлатив») с 44 основами (breda 'широкий', glada 'рад', kalla 'холодный', korta 'короткий', lätta 'легкий', mörka 'темный', närma 'близкий', populära, stör 'большой', svaga 'слабый', svåra 'тяжелый', säkra 'уверенный', troliga 'вероятный', viktiga 'важный', vilda 'дикий', etc.).

Примером случайной протопарадигмы служит пара -erg/-org (обломок основ berg и borg).

Что касается префиксов, то шведская система превербов (an-, av-, be-, in-, med-, om-, på-, upp-, ut-) является единым (хотя и рыхлым) кластером. Как и в немецком языке, максимальное значение S находим в протопарадигме с отрицательным префиксом /o-, 0-/.

2.7. Итальянский язык

Для итальянского языка был взят текст «Декамерона» Боккаччо. Небольшой объем корпуса (260 тыс. словоупотреблений) привел к сравнительно невысоким значениям S.

104	a	o	57	a	i	52	uta	ute
76	i	o	57	osa	oso	52	uta	uti
75	e	i	56	iare	iato	51	are	ato
69	a	e	55	0	no	51	ta	to

³⁸ Это ограничение проявилось и в загадочной паре /-m, -ter/, за которой скрывается /-ism, -ister/ (с 17 основами alkohol, ego, extrem, femin, fundamental, human, ideal, kapital, kommun, naz, nazional, ras, sad, satan, sozial, tur, terror).

69	n	no	53	0	ro	49	0	si
61	uta	uto	53	e	o	48	ar	are
59	ltre	ltro	55	vole	voli	48	ata	ato
						48	e	ono

Сразу же виден «адъективный» кластер из четырех аффиксов.

	a	e	i	o
a		69	57	104
e	69		75	53
i	57	75		76
o	104	53	76	

Эксклюзивная парадигма включает 79 основ (divin, futur, gross, matur, nostr, pruv, quell, quest, timid, tropp, vostr; amoro, disidero, noio, 'скучный', pauro, 'испуганный', pieto, 'согласительный', prezio, 'ценный', valoro). Громадное большинство этих основ кончаются на t – aspettat 'ожидаемый', cadut 'павший', congiunt 'соединенный', legat 'связанный', scelerat 'злодейский', sconsolat 'безутешный', smarrit 'растерянный', stolt 'глупый'. Очень характерны также основы с исходом на issim- (bellissim, bianchissim, carissim, contentissim, grandissim, gravissim, grossissim, lunghissim, nobilissim, onestissim, santissim).

Некоторый намек на «глагольный» кластер показывают связи пяти аффиксов:

	ndo	r	re	sse	va
ndo		35	41	30	37
r	35		-	33	33
re	41	-		40	41
sse	30	33	40		39

Все пять аффиксов присоединяются к 49 основам (aggiugne 'добавить', canta, cerca, conosce 'знать', crede 'верить', dubita, entra, estima 'ценить', intende, manca 'недоставать', manda 'послать', mangia 'есть', marita, opera, parla, passa, pensa, piace 'нравиться', piagne 'плакать', porta, prega 'умолять', prende 'брать', presta 'дать взаймы', richiede 'требовать', ricoglie 'собрать', riconosce 'узнать', risponde 'ответить', tocca 'трогать', torna 'возвратиться', trova 'найти').

Более широкий набор десяти аффиксов находим в варианте все той же глагольной парадигмы:

	ando	ar	are	asse	ava	ò	ata	ate	ati	ato
ando		29	36	31	25	24	28	22	24	33
ar	29		48	31	35	35	39	31	31	41
are	36	48		31	41	43	42	33	34	51
asse	31	31	31		26	28	24	20	23	28
ava	25	35	41	26		35	34	28	30	38
ò	24	35	43	28	35		38	29	32	41
ata	28	39	42	24	34	38		33	38	48
ate	22	31	33	20	28	29	33		28	29
ati	24	31	34	23	30	32	38	28		42
ato	33	41	51	28	38	41	48	29	42	

Все десять аффиксов соединяются с тремя основами aspetta 'ждать', cominci 'начать', raccont 'рассказать'; по 9 аффиксов у cavalc 'ехать верхом', commend 'одобрить', confess 'признавать', giudic 'судить', lasci 'оставить', ragion 'рассуждать', ricord 'помнить', ritorn 'возвратиться', ritrova 'отыскать', trapass 'переступить'. По 8 аффиксов имеют 22 основы (adoper 'употребить', chiam 'звать', comanda 'приказывать', dimor 'проживать', domand 'спросить', guard 'смотреть', procacci 'достать', reputa 'полагать', riguard 'смотреть', sequit 'следовать за', etc.).

Парадигма из трех аффиксов /-0, -la, -lo/ как эксклюзивная включает всего четыре основы donzel 'вести себя по-девичьи', minacciando 'грозя', riconoscendo, riconosciuto 'признавать'; но есть около 70 основ, у которых эта комбинация

сочетается с какими-то другими аффиксами (aiutar 'помочь', aprir 'открыть', cacciar 'гнать', dover 'быть должным', metter 'класть', poter 'мочь', prender 'брать', saper 'знать').

Вырисовывается более широкая парадигма:

	0	gli	la	le	lo	mi	si	vi
0		34	33	22	32	20	29	20
gli	34		24	31	25	20	23	23
la	33	24		23	41		26	22
le	22	31	23		21			
lo	32	25	41	21			27	24
mi	20	20					25	20
si	29	23	26		27	25		25
vi	20	23	22		24	20	24	

По 8 аффиксов имеют 10 основ (avendo 'имея', mostrando 'показывая', tenendo 'держая', trovando 'находя', udendo 'слыша', etc.); по 7 аффиксов у baciar 'целовать', veder 'видеть', voler 'хотеть', по 6 – у dando 'давая', dicendo 'говоря', essendo 'будучи', sentendo 'чувствуя', uccider 'убить'; по 5 – у amando 'любя', menar 'водить', piacer 'нравиться', stando 'становиться', tirando 'тянуть', toccar 'трогать', togliendo 'снимая'.

Среди префиксов сразу же обнаруживается эксклюзивная трехчленная парадигма /0-, d'-, l'/ со средним значением S=83 охватывает 31 основу abito, allegrezza, altezza, altre, altrui, amaritudine, amista, amore, amoroso, andarsi, angoscia, animo, antiche, ardire, ascoltare, asino, aspettare, attendere, autorità, avarizia, averle, avervi, avvedimento; erminia; oltraggio, ombra, onore, orazione, ordine, oriente; ucciderlo (еще 18 основ сочетаются еще с какими-то префиксами). Здесь мы находим полную аналогию с французским языком.

Второй кластер охватывает пять префиксов /a'-, co'-, da'-, de'-, e'-/ , с меньшим основанием к этому кластеру примыкает ne'-. Единственная эксклюзивная парадигма представлена одной основой (giovani), еще две основы комбинируются с пятью префиксами (quali, compagni), по 4 префикса имеют fatti, fratelli, frati, gentili, parenti, prieghi, tafani, vostri.

	a'-	co'-	da'-	de'-	e'-	ne-
a'-		40	64	90	45	32
co'-	40		35	63	22	22
da'-	64	35		44	25	12
de'-	90	63	44		43	31
e'-	45	22	25	43		-
ne'-	32	22	12	31	-	

Некоторым гибридом двух парадигм служит набор /al-, dal-, del-, nel-/. Поскольку все соответствующие основы начинаются на l', четыре префикса могут быть записаны как /all'-, dall'-, dell'-, nell'/. Полный набор префиксов находим у основ altra/e/o, arca, oste, una, uno; по три аффикса находим у abate, altezza, altrui, anime, animo, arme, arte, ira, isola, ordine, opera.

2.8. Латинский язык

Корпусом для проверки алгоритмов здесь послужили тексты Цицерона (1,1 млн. словоупотреблений). Значения S в этом случае чрезвычайно высоки.

439	m	s	203	sima	sime	131	0	ne
400	ne	nem	198	ses	set	131	sem	sent
332	simi	simo	192	0	ur	123	mus	t
296	0	m	184	sent	set	121	a	um
294	nem	nes	166	sem	set	121	e	m0
267	nem	nis	157	se	ses	121	nes	num
267	sime	simo	153	em	es	120	onem	onis
262	0	que	153	sem	ses	118	m	t

258	sima	simi	149	se	sent	116	em	ibus
258	sima	simo	146	m	nt	116	um	us
250	ne	nes	141	ni	nis	115	o	um
241	0	s	141	se	sem	114	nis	num
238	ne	nis	140	ine	inem	114	set	ti
237	se	set	137	one	onem	114	sse	sset
209	nt	t	137	sent	ses	114	0	t
208	sime	simi	135	m	mque	113	ne	ni
204	nes	nis	134	nem	ni			

Прежде чем продвигаться вниз по ранжированному списку S, обратим внимание на две протопарадигмы, практически не входящие в будущие кластеры. Речь идет о паре /-0, -que/ (с ее вариантом /-m, -mque/), присоединяющейся более чем к 800 основам с самыми разными хвостами. Похожа на нее протопарадигма /-0, -ne/ с 110 основами. Есть 28 случаев, где -que и -ne встречаются при одних и тех же основах. Содержательный анализ сразу же квалифицирует их как энклитические частицы, соответствующие русским *и* и *ли*.

Первой возникает агрегированная парадигма с шестью аффиксами:

	-ne	-nem	-nes	-ni	-nis	-num
-ne		400	250	113	238	93
-nem	400		294	134	267	105
-nes	250	294		112	204	121
-ni	113	134	112		141	110
-nis	238	267	204	141		114

Все 6 аффиксов появляются у 22 основ (*cicero, condicio, contio, latro, legio, oratio, praeco, ratio, religio, scipio, sermo, etc.*), по пяти аффиксов имеют 26 основ (*accusatio, affectio, ambitio, caepio, carbo, commu, divinatio, regio, translatio, virgi, etc.*). Подавляющее большинство таких основ имеют исход на *o-*. Ясно, что очень высокие значения S будут наблюдаться и у аффиксов *-one, -onem, -ones, -oni, -onis, -onum*.

Второй агрегированной парадигмой оказался набор /-sima, -sime, -simi, -simo/ (поскольку все основы имеют исход на *is-*, ниже эти аффиксы даются в виде *-issima, etc.*):

	-issima	-issime	-issimi	-issimo
-issima		203	258	258
-issime	203		208	267
-issimi	258	208		332
-issimo	258	267	332	

Все четыре аффикса зафиксированы у 14 основ (*acerb, amant, amic, coniunct, crudel, grav, honest, improb, iust, ornat, prudent, sanct, suav, turp*). По три аффикса имеют еще 24 основы – *acut, antiqu, audac, constant, diligent, doct, familiar, fidel, florent, foed, fort, frequent, grat, human, impur, inimic, iniqu, iucund, nobil, parat, praestant, scelerat, sever, tenui*.

При дальнейшем движении вниз по ранжированному списку постепенно вырисовывается «адъективная» парадигма. Уже в приведенном выше начале списка видны две протопарадигмы (по 60 основ): /-a, -um/ S=121 и /-o, -um/ S=115; здесь же видна пара /-um, -us/ с S=116. Чуть ниже (S=106) видна пара /-i, -um/. Все пять аффиксов зафиксированы у трех сотен основ, например, *abieci, aegypt, alban, apert, campan, carissim, clarissim, consult, consumpt, content, copios, cotidian, decor, delect, dissolut, doctissim, domin, erudit, exanimat, exigu, facet, faciend, faust, ferend, fidelissim, furios, gratissim, habit, honest, ignot, improb, impur, inaudit, incens, incert, inclus, indign, ingenios, infinit, ingenu, inhuman, inimic, iniqu, iniust, intim, iucund, legend, leontin, moderat, nefari, neglegend, nobilissim, nunti, occis, occupat, oppress, ornat, ornatissim, otios, perdit, perfect, perfunct, physis, posit, postrem, praeposit, primari, pristin, product, profect, prompt, propinqu, propri, prudentissim, pulcherrim, quiet, religios, remot, restitut, retent, rustic, scelerat, script, sanct, sever, socratic, sollicit, superb, syracusan, temperat, timend, tribut, turbulent, tusculan, urban, usitat, vitios, vituperand*. В этот перечень не

включены основы, сочетающиеся одновременно с нулевым аффиксом (например, advocat, appetit, armat, audit, dedit, legat, liberat, privat, quint).

Двигаясь вниз по списку, мы обнаружим новые аффиксы, связанные с -um и с -a: -ae, -am, -arum, -is, -orum, -os. По мере расширения парадигмы она все больше теряет специфически «адъективный» характер, превращаясь в широкую «именную» парадигму. Внутренние связи в парах этих аффиксов отражены в следующей матрице.

	a	ae	am	arum	as	i	is	o	orum	os	um	us
a		84	83	24	48	67	84	79	44	65	121	81
ae	84		84	46	60	44	57	47	22	52	59	52
am	83	84		25	66	52	61	64	23	52	65	58
arum	24	46	25		39	11	29	12	20	14	13	13
as	48	60	66	39		35	52	50	16	34	35	31
i	67	44	52	11	35		99	77	44	70	106	75
is	84	57	61	29	52	99		75	51	65	92	50
o	79	47	64	12	50	77	75		47	59	115	84
orum	44	22	23	20	16	44	51	47		55	51	39
os	65	52	52	14	34	70	65	59	55		84	80
um	121	59	65	13	35	106	92	115	51	84		116
us	81	52	58	13	31	75	50	84	39	80	116	

Внутренние контрасты, которые могли бы соответствовать традиционным субстантивным парадигмам 1 и 2 склонения, практически незаметны в этой матрице. Едва проглядывают повышение S в группе первых пяти аффиксов. По числу аффиксов именные основы распределяются так:

- 12 - domestic, public;
- 11 - alien, antiqu, patern, plurim, praeterit, privat, proxim, reliqu, secund, univers;
- 10 - acerb, aetern, contrari, corrupt, debit, decret, divin, extern, futur, human, iniqu, latin, maxim, necessari, obscur, optim, perfect, probat, quant, relict, roman, urban;
- 9 - absolut, armat, attic, barbar, comparat, conduct, domin, graec, infirm, iudicat, maritim, mortu, nostr, recondit, terti, ultim, vestr, violat;
- 8 - advocat, audit, african, clodi, dialectic, habit, magistr, minim, miser, nocturn, perpetu, tacit, timid, tractat;
- 7 - ambgu, conservat, foederat, german, gratios, legitim, liber, quint, rhetoric, ridicul, tyrann, vicin.

В тех случаях, когда число аффиксов снижается дальше, семантический анализ выделил бы и явно субстантивные основы, сгруппированные ниже в соответствии с типом склонения и с грамматическим родом.

- 6 - decim, elect, exact, gener, horrid, invict, legat, numer, quart;
- f) coloni, column, discordi, epistul, fortun, histori, iniuri, insul, litter, pecuni, sententi, tabell, vigili;
- n) argent, consili, exempl, imperi, monument, oracul, ornament, peccat, praemi, testament, testimoni, vocabul;
- m) gladi, praefect;
- 5 - concept, diurn, educat, neutr, troian;
- f) coron, doctrin, hispani, lacrim, laetiti, pirat, victori;
- n) artificii, auguri, edict, elect, firmament, ingeni, oppid, praesidi, preti, sesterti, somni, spatii, studi,
- m) colon, medic, patricii.

Вернемся, однако, к началу ранжированного списка. Очень быстро формируется шестичленная «перфектная» парадигма /-se, -sem, -sent, -set, -ti/; поскольку соответствующие основы неизменно оканчиваются на is-, наши аффиксы могут записываться как -isse, -issem, -issent, -isset, -isti.

	-se	-sem	-sent	-ses	-set	-ti
-se		141	149	157	237	94
-sem	141		131	153	166	71
-sent	149	131		137	184	57

-ses	157	153	137		198	70
-set	237	166	184	198		114
-ti	94	71	57	70	114	

Эксклюзивная парадигма объединяют шесть основ: *maluis, respondis, retinuis, scripsis, suscepis, tradidis*; еще десять основ сочетаются одновременно с какими-то другими аффиксами: *accepis, contulis, gessis, habuis, meminis, perdidis, potuis, rettulis, sustulis, voluis*. Оставаясь в рамках четырехбуквенного ограничения на длину аффикса, проверим наше предположение об *-is-* как части аффикса. Тогда возникнет еще один вариант все той же перфектной парадигмы:

	-isse	-isti	-erat	-erim	-eris	-erit	-ero
-isse		94	79	60	73	102	43
-isti	94		64	48	65	86	40
-erat	79	63		47	62	86	40
-erim	60	48	47		45	53	43
-eris	73	65	62	45		74	37
-erit	102	86	86	53	74		37
-ero	43	40	40	43	37	37	

Все 7 аффиксов зафиксированы у трех основ – *cognovi, concess, scrips*; по 6 аффиксов имеют *constitu, contemps, defend, discess, ostend, perven, reliqu, respond, suscep*.

Пять аффиксов */-0, -m, -nt, -s, -t/* образуют ядро «глагольной» парадигмы. В то время как *-m* и *-s* часто фигурируют как исход «именных» аффиксов (*-am, -as, -os, -um, -us*), *-t* и *-nt* характерны для данного набора. Если, спускаясь по списку ко все более низким значениям *S*, следить за связями *-t* и *-nt*, мы обнаружим следующие пары с новыми аффиксами: */-t, -mus/ S=123, /-nt -mus/ S=111, /-t, -tur/ S=110, /-nt, -tur/ S=85, /-t, -ntur/ S=75, /-t, -sse/ S=68, /-t, -ntur/ S=64, /-t, -sset/ S=63, /-t, -sti/ S=63, /-t, -re/ S=53, /-nt, -bat/ S=52, /-t, -tis/ S=47, /-nt, -ret/ S=47, /-nt, -re/ S=46, /-t, -ret/ S=45, /-t, -bat/ S=42, /-nt, -tis/ S=40*.

Теперь «глагольная» парадигма расширилась до 12 аффиксов.

	-0	-s	-m	-t	-nt	-mus	-tur	-ntur	-re	-ret	-bat	-tis
-0		(241)	(296)	114	59	39	45	23	16	11	7	13
-s	(241)		(439)	103	85	69	54	28	20	17	19	64
-m	(296)	(439)		118	146	74	55	46	6			26
-t	114	103	118		209	123	110	64	53	45	42	47
-nt	59	85	146	209		111	85	75	46	47	52	40
-mus	39	69	74	123	111		66	41	24	25	35	47
-tur	45	54	55	110	85	66		91	42	27	26	38
-ntur	23	28	46	64	75	41	81		36	27	28	22
-re	16	20	6	53	46	24	42	36		87	55	18
-ret	11	17		45	47	25	27	27	87		70	18
-bat	7	19		42	52	35	26	28	55	70		14
-tis	13	64	26	47	40	47	38	22	18	18	14	

Лишь одна основа (*accusa*) сочетается с 11 аффиксами, 10 аффиксов имеют 4 основы (*intellege, nomina, retine, sustine*); по 9 аффиксов находим у 29 основ (*adhibe, appella, audire, cognoscere, commemora, conserva, debere, delecta, demonstra, desidera, dicere, disputa, dubita, excita, existima, habere, ignora, impera, labora, lauda, ostende, proba, putaba, putare, relinque, require, significa, videa, videba*; по 8 аффиксов у 40 основ – *adiuva, antepona, audia, cena, cogita, commove, conceda, confera, confirma, defende, diceba, existime, expecta, gaude, incita, indica, intellega, intellegere, iudicare, narra, numera, observa, obside, petere, posside, postula, probare, probe, pugna, quaera, quaere, reddere, refera, responde, specta, spera, supera, tenea, timere, venire*.

Последний кластер, различимый в верхней части ранжированного списка, включает шесть аффиксов (*-e, -em, -es, -ia, -ibus, -ium*), к которым примыкают еще два (*-i* и *-is*), уже вошедшие в именные парадигмы.

	-e	-em	-es	-ia	-ibus	-ium	-i	-is
-e		108	83	23	71	22	81	86
-em	108		153	53	111	55	77	93
-es	83	153		49	111	53	58	70
-ia	23	53	49		43	67	29	26
-ibus	71	111	111	43		56	40	60
-ium	22	55	53	67	56		32	30
-i	81	77	58	29	40	32		(99)
-is	86	93	70	26	60	30	(99)	

Как видим, пять аффиксов из этого набора (-e, -em, -es, -i, -is) уже появлялись как хвосты аффиксов первой именной парадигмы латинского языка. По-видимому, новая парадигма есть обобщение той первой парадигмы. В порядке уменьшения набора аффиксов группы основ располагаются следующим образом:

8 – accusator, animant, caelest, censor, dissimil, excellent, forens, gladiator, immortal, imperator, mortal, natural, orator, popular, praetor, princip, provincial, sacerdot, sapient, senator;

7 – absent, action, admirabil, adulescent, aequal, agrest, audient, augur, calamitat, ciceron, civitat, cogitation, cohort, commun, condicion, coniug, coniunction, consular, contention, cupiditat, defension, definition, disputation, exercitation, familiar, fidel, hereditat, histrionm illustr, innocent, insign, inutil, largition, latron, legion, locuplet, mancip, mediocr, municip, opinion, oration, parent, pontific, potestat, praedon, praesent, probabil, quaestion, quaestor, ration, religion, salutar, scipion, sermon, simil, societate, suspicion, tempestat, utilitat, victor, voluntat, voluptat;

6 – aedil, animal, arator, ardent, auctor, auctoritat, auditor, caepion, civil, class, client, comit, conclusion, consolation, contion, custod, defensor, delectation, dicent, difficil, dignitat, dissension, dissentient, efficient, errant, facil, familiarit, fatal, flamin, grand, gratulation, humil, imman, incolum, incredibil, inferior, liberal, libidin, milit, mirabil, muliebr, mulier, nobil, nocent, paction, petent, petition, praecon, procurator, pueril, rhetor, rogation, sedent, singular, stabil, superior, tolerabil, valent, vectigal, venal, virtut, vulgat.

Лишь при S=65 обнаруживаются новая пара протоаффиксов /-ior, -ius/ с 83 основами – acerb, accommodat, admirabil, alien, angust, antiqu, apert, ardent, asper, atroc, audac, callid, celer, commod, coniunct, constant, copios, crebr, crudel, cupid, deform, difficil, diligent, divin, elegant, erudit, explorat, facil, felic, fortunat, frequent, grand, honest, ieiun, improb, impudent, impur, incitat, inconstant, infer, infirm, inhuman, inimic, inlustr, iucund, laudabil, liber, liberal, miser, moderat, molest, munit, neglegent, nobil, obscur, occult, optat, ornat, parat, poster, praestant, press, prudent, pulchr, remiss, robust, sanct, sapient, scelerat, splendid, studios, subtil, super, taetr, temperat, timid, tolerabil, trist, valent, vehement. Все эти основы сочетаются с характерными аффиксами (исключение – две основы deter и inter). Весь набор можно охарактеризовать как «компаративное» ответвление «адъективной» парадигмы.

По сравнению с суффиксальными концами латинские превербы характеризуются низкими значениями S и крайней рыхлостью всей системы.

	0	ad	con	de	in	ob	per	pro	re
0		52	56	46	78	10	41	34	39
ad	52		45	24	21	9	32	38	21
con	56	45		34	56	35	39	33	47
de	46	24	34		24	15	19	33	31
in	78	21	56	24		11	19	23	27
ob	10	9	35	15	11		14	4	23
per	41	32	39	19	19	14		31	27
pro	34	38	33	33	23	4	31		27
re	39	21	47	31	27	23	27	27	

По числу присоединяемых аффиксов располагаются группы основ:

9 – ducere

8 – ferre, tulit

7 – ditum, fecit, latum, scribere, sequi

6 – esset, iecit, missum, vertit

5 – cessit, gredi, ponere, posita, spexit, stitit, venisset, vocatio.

Вторым латинским корпусом стал текст Вульгаты (500 тыс. словоупотреблений). При том, что Цицерона и Иеронима разделяют четыре с половиной века, при всех различиях жанров и идеологии мы могли бы ожидать существенных изменений. Однако два ранжированных списка очень похожи.

437	ne	nem	114	ni	nis	93	dant	dent	85	se	set
333	nem	nis	110	set	ti	93	sent	set	84	abit	avit
326	0	que	109	rum	s	92	0	e	84	ant	unt
281	m	s	107	nem	num	92	it	unt	84	at	unt
277	nem	nes	107	one	onis	92	nes	num	84	iens	ient
275	ne	nis	106	enta	ento	90	dent	dite	83	dite	dunt
232	ne	nes	106	em	es	90	tem	tes	83	o	um
179	0	m	105	s	t	89	0	rum	82	erit	it
178	0	ur	103	ili	ilia	88	ne	num	82	m	nt
171	nes	nis	100	em	ibus	88	on	ones	81	dat	det
165	nt	t	99	nis	num	87	i	um	81	em	is
165	0	s	99	onem	ones	87	sem	set	81	scet	scit
138	one	onem	98	es	ibus	87	tiae	tiam	80	abo	avi
130	nem	ni	97	mus	t	86	cite	cunt	80	cens	cent
128	onem	onis	96	bus	s	86	e	m	80	cens	cere
114	ne	ni	94	a	um	86	scas	sces	80	ient	iunt

Конечно, какие-то различия есть. Например, протопарадигма /-0, -que/ передвинулась с восьмого места на третье и охватывает теперь 973 основы (при вдвое меньшем корпусе). Нет ли здесь влияния еврейского оригинала с его wa-?

Значения S у второй агрегированной парадигмы уменьшились в 5-6 раз, хотя она по-прежнему сохраняется как целое.

Как и в корпусе Цицерона, первой открывается шестичленная парадигма /-ne, -nem, -nes, -ni, -nis, -num/ (все эти аффиксы сочетаются с основами cogitatio, cognatio, generatio, natio, regio, sermo, ultio «мечь»).

Хорошо представлен пятичленный набор /-a, -i, -o, -um, -us/ преимущественно «адъективной» парадигмы, полностью представленный у accept, aetern, alien, antiqu, apert, benedict, chanane, concept, contrari, corrupt, decept, decim, desert, dilect, divers, divis, domin, elect, erect, excels, extrane, extrem, famul, finit, futur, habit, hebrae, human, incens, incert, infern, incicumcis, infirm, ingress, inimic, iniqu, iniust, inmund, maledict, maxim, modic, mortu, novissim, numer, nunti, optim, peregrin, perfect, pervers, pessim, philisthe, plurim, posit, pretios, primogenit, quant, quart, quint, relict, reliqu, robust, sanct, secund, sempitern, stult, superb, terti, timid, univers, vicin, violent, etc. Всего 106 основ, а не 300, как у Цицерона.

Хорошо сохраняются и остальные парадигмы. Единственное исключение – исчезновение «компаративной» парадигмы. При шести основах (ecels, exter, infer, inter, sanct, sublim) с аффиксами -ior и -ius имеют S=18, что недостаточно для включения в статистически значимый список парадигм.

Что касается провербов, то их набор несколько изменился, а значения S – повысились.

	0	ad	con	de	ex	in	per	pro	re
0		62	76	48	41	88	46	37	49
ad	62		49	27	18	27	17	25	33
con	76	49		46	37	45	35	23	62
de	48	27	46		10	37	19	29	44
ex	41	18	37	10		29	20	8	25
in	88	27	45	37	29		22	36	37
per	46	17	35	19	20	22		15	25
pro	37	25	23	29	8	36	15		30
re	49	33	62	44	25	37	25	30	

По 9 префиксов имеет *posuit*; по 7 – *cepit, ducet, ponam, tracet*; по 6 – *cessit, latus, mittere, secuti, tendit*; по 5 – *clusit, ferre, gressus, ierat, positus, spicis, surgentes, venient, versi, versus, vocare*.

2.9. Поиски парадигм – краткие итоги

Подведем итог попыткам выделения парадигм, описанным во второй части книги. В п. 2.1. процесс выделения суффиксальных и префиксальных парадигм осуществлялся с некоторой оглядкой на частоту буквенных цепочек, но постепенно центр тяжести процедуры переносился на выявление регулярностей в комбинаторике возникающих аффиксов и основ независимо от их частотных характеристик. В пп. 2.2–2.8 частота учитывается ровно один раз – при определении порога для выделения достаточного большого списка графических слов³⁹. Для Корпуса русской прозы список включает 96 тысяч слов; корпус «Войны и мира» – 22 тысяч слов, корпус «Независимой газеты» – 136 тысяч слов; корпус газеты «The Times» включает 70 тысяч слов; корпус Гофмана – 24 тысячи слов; корпус «Süddeutsche Zeitung» – 108 тысяч слов; «Göteborg Posten» – 78 тысяч слов; корпус Цицерона – 43 тысяч слов.

Общей базой для расчета математических ожиданий у пар протоаффиксов становится теперь число основ, сочетающихся с аффиксами. Значения *S* подсчитываются для пар протоаффиксов, которые могут складываться в какие-то кластеры – агрегированные парадигмы. Конечно, появляющиеся парадигмы интерпретируются как ядра будущих окончательных парадигм, чей состав будет окончательно уточнен на минимальном интервале.

В русском языке лучше и раньше всего формируются глагольные парадигмы (включая причастия и некоторые деепричастия), затем – адъективные парадигмы (включая компаративы и наречия на *-о*) и затем (и с большим трудом) – субстантивные парадигмы. Статистическая процедура выделения аффиксов и основана на представлении о преимущественно агглютинирующем характере присоединения аффиксов к основам. Лишь изредка при анализе пар протоаффиксов удается заметить какие-то приметы регулярных морфонологических явлений (например, беглых гласных, чередований типа *г/ж, к/ч*).

В английском языке раньше всего выделяется субстантивная парадигма, затем – глагольная парадигма, и наконец – адъективная.

Последовательность выделения парадигм во французском языке напоминает сценарий русского языка. Специфической чертой французского и итальянского языков можно считать появление префиксов-проклитик с исходом на *'*. В обоих романских языках субстантивы не были опознаны как самостоятельные типы.

В немецком языке выявились адъективные и глагольные парадигмы, но субстантивные парадигмы остались неопознанными.

В шведском языке довольно легко выделяются типы парадигм, ассоциируемых с тремя традиционными частями речи. Субстантивные парадигмы опознаются из-за постпозитивного артикля – специфической черты скандинавских языков.

Как и следовало ожидать в случае древних индоевропейских языков скорее должна выделиться единая именная парадигма, нежели противопоставленные друг другу субстантивные и адъективные парадигмы. С другой стороны, вполне ожидаемы различия продуктивного (традиционных первого и второго) склонения и традиционного третьего склонения, особая перфектная парадигма.

В пп. 2.1 и 2.2 уже указывались возможности уточнения парадигм в пределах микроинтервала (то есть – в пределах графического слова). Основная надежда, однако, связана с переходом к минимальному интервалу.

Оставаясь в рамках микроинтервала, обратимся к некоторым примерам разрешения аффиксальной омонимии. Первым примером послужит аффикс *-ую*, зафиксированный в наборах трех широких адъективных парадигмах ТЯЖЕЛ-ый, ХОРОШ-ий и ВЫСОК-ий. Но *-ую* может встречаться при глагольных основах с совсем другими наборами аффиксов. Характерна парадигма ЧУВСТВ-овал с набором */-овал, -уем, -ует, -уете, -уй, -уют, -уя/*. Кроме основ с исходом на *ств-* здесь представлены,

³⁹ Такой порог не вводился в следующих случаях: корпус Диккенса, где список включает более 50 тыс. графических слов, корпус Шекспира со списком в 28 тыс. графических слов, корпус французских романов – 86 тысяч слов, корпус «Декамерона» – 19 тысяч слов, корпус Вульгаты – 43 тысяч слов.

например, *бесед-, совет-, жертв-, арест-, пожал-, по/растолк-, досад-, рекоменд-, завид-, проповед-, негод-, претенд-, послед-, требов-*.

Аффикс *-у* зафиксирован в трех важнейших субстантивных парадигмах ОБРАЗ, ЖЕНИХ и ПРИЧИН-а, но он же соединяется более чем с сотней глагольных основ с разными наборами аффиксов (*высек-, долож-, изорв-, навещ-, позов-, полеж-, приглаш-, напиш-, расшиб-, сберег-, хлопоч-, разбер-, взглян-, отдохн-, возвращ-, возъм-, выраст-, выпущ-, выруч-, вытащ-, заключ-, закрич-, замеч-, замолч-, огорч-* и т. п.). Таким образом, выясняется, что и в глагольных парадигмах присутствует аффикс *-у*.

После того как выявлены регулярные агрегированные парадигмы, можно вернуться к протопарадигмам для поиска буквенных чередований.

г/ж — *заслу-г/жил, предло-г/жил, трево-г/жи, интри-г/жки, подвер-г/жен, прибе-г/жит, коври-ги/жки;*

г/ч — *запря-г/чъ, дости-г/чъ, стере-ги/чъ, приле-г/чъ, сбере-г/чъ;*

д/ж — *перев-д/жу, переез-д/жать, перехо-д/жу, пропа-ду/жа, госпо-да/жа, ненави-дит/жу, предупре-дил/жу, разбу-дил/жу, рассер-дил/жен;*

з/ж — *вообра-зил/жал, возра-зил/жал, прибли-зил/жать, расска-зал/жет, сконфу-зил/жен, развя-зал/жет;*

к/ч — *земля-к/чок, стари-к/чо, колпа-к/чок, червя-к/чка, свече-к/чка, дожди-к/чек, плато-к/чек, челове-к/чек, голубо-к/чка, пересе-к/чъ, привле-к/чъ, замол-к/чал, праздни-к/чный, практи-ка/чный, полити-ка/чный;*

к/ц — *сплетни-к/ца, учени-к/ца, мучени-к/ца, чиновни-к/ца, помещи-к/ца. любовни-к/ца, мельни-к/ца.*

Дистрибутивно-статистический анализ в минимальном интервале на этом не заканчивается. На каком-то этапе исследования можно вернуться к учету частот графических слов и их групп. При этом может проявиться внутренняя структура агрегированных парадигм (как формальная, так и предположительно — семантическая). Обратимся для демонстрации к анализу двух субстантивных парадигм ОБРАЗ и ЖЕНИХ с идентифицируемым набором аффиксов /0, а, ам, ами, ах, е, ов, ом, у, ы/и/. Отберем самые частые основы, сочетающиеся с аффиксами этого набора (включим сюда и трехбуквенные основы):

Таблица 2.2

	0	а	ам	ами	ах	е	ов	ом	у	ы/и	
глаз	2994	11056	458	4451	2647	14	5	296	298	-	=22219
дом	4657	7085	293	64	292	3936	495	463	1787	121	=18893
бог	6966	2687	25	8	66	57	58	967	2598	58	=13492
голос	4738	1432	25	99	5	551	421	4590	225	-	=12086
год	3044	3742	73	215	189	26	522	184	1689	915	=10599
вид	2411	385	13	39	204	2041	112	1367	1272	158	=8002
вечер	2437	1767	411	27	91	179	86	2432	555	-	=7985
город	2744	1711	99	25	174	1833	179	224	703	-	=7692
вопрос	2984	568	77	239	94	171	351	465	144	1056	=6149
народ	3013	1119	19	23	4	352	181	432	831	138	=6112
взгляд	2192	748	32	128	69	428	191	1407	114	414	=5723
граф	3425	1252	9	6	4	72	38	249	490	59	=5604
двор	1732	1107	94	25	63	1732	126	80	449	91	=5499
мужик	1751	654	265	195	24	15	758	121	226	1296	=5305
генерал	2516	886	25	56	11	28	192	229	369	215	=4527
мир	1498	954	4	1	4	896	28	284	516	20	=4205
дух	1172	1219	4	90	8	446	114	636	412	72	=4183
лес	1168	901	131	62	227	38	181	232	1176	48	=4164
доктор	2135	784	15	25	1	19	88	282	204	-	=3553
круг	443	258	4	30	36	44	17	2012	332	93	=3271
мальчик	1651	761	32	35	10	42	163	179	186	208	=3265
воздух	1102	264	-	-	1	1087	8	219	215	6	=2902
диван	934	468	12	17	24	872	28	99	78	40	=2572
кабинет	1212	458	1	5	2	709	13	40	86	6	=2532
карман	839	729	34	8	74	497	48	15	41	209	=2494
грех	1267	426	61	15	61	45	82	120	19	322	=2418
закон	675	286	153	72	54	91	260	143	357	327	=2418

век	1056	413	3	79	15	72	200	23	262	258	=2381
звук	516	213	63	119	96	70	163	102	50	933	=2325
крик	846	101	14	155	25	21	84	418	84	567	=2315
дурак	1173	226	36	50	45	6	133	204	71	223	=2167
губернатор	908	513	11	13	5	29	35	137	356	42	=2049
зуб	109	51	46	404	106	2	181	17	4	1088	=2008

Всего проанализировано 188 основ, получены следующие данные о доле каждого аффикса в общей сумме.

0	а	ам	ами	ах	е	ов	ом	у	ы/и
37.7%	19.0%	1.3%	1.9%	1.4%	8.0%	4.9%	10.3%	8.3%	7.2%

Превратим количественные данные Таблицы 2.2 в качественные, превратив в плюсы превышение усредненной доли аффикса у всех 188 основ. Таблица примет такой вид:

	0	а	ам	ами	ах	е	ов	ом	у	ы/и
бог	+	+							+	
век	+			+			+		+	+
вечер		+	+					+		
взгляд	+			+				+		+
генерал	+	+								
год		+		+	+				+	+
зуб			+	+	+		+			+

Теперь к столбцам новой таблицы применим формулу расчета коэффициента связи, уже использованного в п. 1.4.4. В результате обнаружился кластер пяти аффиксов со следующими значениями Q.

	-ам	-ами	-ах	-ов	-ы/и
-ам		+47	+85	+60	+57
-ами	+47		0	+69	+82
-ах	+85	0		0	+30
-ов	+60	+69	0		+72
-ы/и	+57	+82	+30	+72	

К числу основ, часто сочетающихся с этими пятью аффиксами (т. е. имеющих плюс в соответствующих столбцах), относятся билет, вывод, гусар, дурак, закон, звук, зуб, идеал, интерес, куст, литератор, наряд. По четыре аффикса из этих пяти неслучайно часто соединяются с основами актер, амбар, арестант, артист, башмак, бокал, бурлак, визит, висок, воз, воин, враг, допрос, доход, жандарм, жид, жилец, инстинкт, источник, казак, каприз, луг, магазин, матрос, мешок, миллион, молодец, монах, мошенник, мужик, намек, наследник. Таким образом, семантика множественного числа оказалась настолько мощной, что она проявилась в высоких значениях Q. Высокую связь обнаружили аффиксы -ах и -е (Q=+70). Правда, к этой паре аффиксов обычно присоединяется еще какой-то аффикс:

-а	-е	-ах	акт, балкон, городок, комитет;
-0	-е	-ах	жилет;
-ы	-е	-ах	кафтан, мундир;
-а	-ы	-е -ах	карман;
-0	-ы	-е -ах	костюм.

Любопытна семантическая специфика основ, сгруппировавшихся вокруг пары -ах и -е. Некоторые семантические инварианты могут быть характерны для основ, неслучайно часто сочетающихся с той или иной комбинацией аффиксов в нашем наборе, ср. /-0, -а, -у/ - бог, граф, губернатор, дьявол, командир, майор, но также денщик и кучер; /-0, -а, -у/ автор, генерал, дворник, доктор, капитан, мальчуган, наполеон; /-0, -а, -ов/ дьячок, лоцман, мальчик, мастер.

Некоторые клетки таблицы 2.2 остались незаполненными, и в случае очень частых слов это пустое место статистически значимо, ср. глаз (S=-38), голос (S=-29), вечер (S=-23), город (S=-23), доктор (S=-15). Интерпретация этих четырех

случаев очевидна с позиций семантической лингвистики – формы множественного числа у этих основ образуются с помощью флексии -а. Высокие отрицательные значения S свидетельствуют о грамматических запретах и в некоторых других строках таблицы 2.2, ср. *дом* (S=-33), где зафиксированы формы -ы, уступающие под напором -а. Но отрицательные значения S очень велики и у основ *бог* (S=-29), *мир* (S=-16), *кабинет* (S=-13), где действуют чисто семантические факторы; что касается основы *лес* (S=-13), то здесь проявилась омонимия со словом *леса*.

Кластер /-ам, -ами, -ах/ изначально обладает сильными внутренними связями (среднее значение S=121); этот кластер – составная часть парадигм ОБРАЗ и ЖЕНИХ, но его можно зафиксировать в некоторых других парадигмах. Так, если к парадигме ПРИЧИНА добавить аффиксы -ам, -ами, -ах и -0, они органично сольются с уже выделенным набором /-а, -е, -ой, -ою, -у, -ы/, но сохраняют свою компактность и автономность (среднее значение Q=+70). На этот раз три аффикса плюралиса (-ам, -ами, -ах) соединяются с -0, а не с -ов, как это было в «мужских» парадигмах. Характерные основы *баб, газет, женщин, звезд, карт, картин*.

По-видимому, дальнейшее исследование морфологических структур следует отложить до того момента, когда оно сможет сопрягаться с новой информацией, получаемой на минимальном интервале.

Таблица 3.4

100 самых частых существительных

33010	дело	6362	сын
32634	рука	6325	народ
26118	человек	6247	земля
25913	время	6084	вопрос
22163	глаза	6079	дорога
20731	день	5974	Марья
19720	слово	5880	свет (<i>monde</i>)
19656	голова	5750	вода
17775	люди	5747	гость
17767	лицо	5741	утро
16773	жизнь	5650	дети
15737	дом	5610	слеза
14705	князь	5524	взгляд
14424	отец	5510	двор
14138	место	5509	дочь
12273	деньги	5495	ум
12048	сторона	5476	господин
11962	Бог	5427	мужик
11826	голос	5413	Москва
11791	минута	5411	Анна
11618	комната	5408	Петр
10462	мать	5407	конец
10218	душа	5315	девушка
10131	год	5244	рубль
10057	женщина	4850	граф
9802	час	4821	батюшка
9714	сердце	4811	Петербург
9670	дверь	4811	положение
9153	жена	4729	улица
8959	нога	4680	Николай
8818	мысль	4643	книга
8780	брат	4599	шаг
8553	лет	4582	хозяин
8543	Иван	4567	часть
8399	ночь	4559	генерал
8337	стол	4555	дама
8011	сила	4482	господа
7956	случай	4440	имя
7827	друг	4422	матушка
7800	муж	4417	пора
7754	старик	4393	плечо
7728	вечер	4385	счастье
7713	город	4269	смерть
7688	вид	4233	Алексей
7644	письмо	4219	обед
6922	любовь	4176	деревня
6747	чувство	4157	лес
6634	окно	4157	общество
6444	лошадь	4054	улыбка
6439	разговор	3961	Павел

100 самых частых прилагательных

12024	большой	2449	огромный
11052	новый	2404	лучший
9479	должен	2346	пустой
9125	последний	2325	честный
8627	старый	2294	главный
7955	хороший	2276	красивый
7781	добрый	2265	тяжелый
7134	целый	2250	тихий
7059	молодой	2232	пьяный
6921	русский	2215	довольный
5919	маленький	2127	широкий
5121	великий	2109	важный
5102	белый	2047	бледный
4720	милый	2024	благородный
4711	разный	2021	несчастный
4477	полный	1984	холодный
4415	черный	1952	женский
4268	бедный	1927	человеческий
3907	прочий	1909	глубокий
3749	настоящий	1900	дорогой
3748	прежний	1865	малый
3565	знакомый	1818	легкий
3484	общий	1786	дурной
3474	рад	1774	похожий
3431	высокий	1752	остальной
3416	готов	1726	следующий
3415	странный	1723	спокойный
3379	собственный	1714	высший
3216	умный	1709	обыкновенный
3169	веселый	1704	приятный
3142	красный	1696	мелкий
3050	длинный	1684	больший
3033	сильный	1683	тонкий
2958	особенный	1650	старший
2943	счастливый	1648	слабый
2924	небольшой	1644	серьезный
2912	простой	1622	золотой
2906	родной	1621	военный
2887	темный	1620	строгий
2868	живой	1600	светлый
2851	чужой	1594	одетый
2838	больной	1580	горячий
2795	подобный	1565	ужасный
2788	прекрасный	1556	толстый
2727	богатый	1513	уверен
2621	святой	1491	злой
2597	глупый	1478	мертвый
2575	божий	1469	зеленый
2513	чистый	1442	нежный
2454	внутренний	1426	любезный

100 самых частых глаголов

179941	быть	5893	сесть
60539	говорить	5741	помнить
58142	сказать	5721	увидеть
45244	знать	5605	оставить
36562	мочь	5565	лежать
24048	стать	5533	спать
21451	видеть	5489	спрашивать
20831	хотеть	5407	выходить
19861	думать	5231	давать
18072	спросить	5191	брать
18029	отвечать	5106	получить
17397	идти	5096	принять
16795	любить	5088	начинать
14816	сделать	4987	прибавить
14733	жить	4868	написать
14579	делать	4842	хотеться
13322	смотреть	4800	верить
13060	стоять	4793	забыть
12981	сидеть	4788	оставаться
12825	выйти	4755	ждать
12709	начать	4721	ответить
10792	прийти	4637	подойти
9508	продолжать	4557	поехать
9251	понимать	4482	взглянуть
8947	заметить	4476	держать
8483	ходить	4468	встать
8272	пойти	4464	молчать
8121	просить	4459	смеяться
7969	слышать	4453	пить
7960	остаться	4434	кричать
7772	глядеть	4432	играть
7385	ехать	4377	остановиться
7232	годиться	4355	произнести
7183	приехать	4315	отдать
7020	проговорить	4246	показаться
6836	слушать	4216	умереть
6817	понять	4181	считать
6812	бывать	4176	велеть
6795	писать	4128	стараться
6428	иметь	4037	подать
6414	значить	3998	приходить
6383	читать	3958	рассказывать
6280	бояться	3939	воскликнуть
6220	желать	3937	уходить
6117	чувствовать	3933	уметь
6078	подумать	3907	рассказать
6061	найти	3844	обратиться
6040	посмотреть	3840	служить
5971	пройти	3813	уехать
5928	уйти	3793	успеть