

Исследование лингвистической адекватности моделей FTD жанровой классификации текстов

Павлова А. Ю. (pavlova-alex.u@yandex.ru)
РГГУ, Москва, Россия

In this study a system of functional text dimensions is taken as a basis for investigation of genre diversity. We study the differences in representation of linguistic features in genres, as well as how the same linguistic phenomena functions can vary depending on the text genre. As a result, we confirm the fact that genre characteristics of the material for linguistic research is the factor that must be considered carefully.

Keywords: corpora, genre classification, FTD, linguistic features.

Введение

Предполагается, что в жанрах должно варьироваться представление лингвистических признаков, а также одни и те же языковые явления могут функционировать по-разному, что значит, при проведении любого исследования необходимо принимать во внимание жанровую принадлежность текстов. То есть информация относительно жанровых характеристик может представить собой определяющий фактор для объективности полученных результатов. По этой причине далее будет рассматриваться набор признаков, статистически значимые различия и взаимосвязи в использовании различных единиц, что подтвердит валидность подобного разделения.

В рамках данного исследования в классификации за основу берется система функциональных текстовых размерностей, предложенная С. А. Шаровым (Sharoff, 2018). Каждая размерность — это функциональная категория, которая характеризует цель текста. Например, в аргументативных текстах — это проведение и поддержание дискуссии, в инструктивных - дать совет, помочь, а в нормативно-правовых - установление нормы.

В предыдущих исследованиях со схожими целями (Lagutin, 2015 и Katinskaya, 2016) материал для рассмотрения особенностей FTD был небольшим (примерно 500 текстов), то некоторую специфику не было возможности выявить достоверно.

Материал

Основным материалом для исследования послужил датасет, составленный из 16 тысяч текстов, после отсека неподоходящих для наших целей текстов во время жанровой разметки (полностью отбрасывались автоматически сгенерированные тексты и слишком короткие) осталось: 1275 текстов РИА Новости, 7454 текстов Яндекс Дзен, 6086 текстов Живого Журнала. Данные тексты были размечены независимо парами\тройками разметчиков и все значительные расхождения были далее были также согласованы. Значительными в данном случае считались все случаи, где основным жанром (разметчики присваивают значение 2 для FTD) были выбраны разные функциональные размерности.

В данных источниках очевидна значительно большая неоднородность в плане представленных функциональных размерностей.

РИА			Дзен			ЖЖ		
A8	1213	95,14%	A7	1957	26,25%	A11	2373	38,99%
A13	28	2,20%	A8	1395	18,71%	A1	1226	20,14%
A1	14	1,10%	A1	1275	17,10%	A4	603	9,91%
A16	9	0,71%	A16	821	11,01%	A8	539	8,86%
A20	5	0,39%	A11	697	9,35%	A17	412	6,77%

A14	2	0,16%	A17	543	7,28%	A7	276	4,53%
A12	2	0,16%	A4	366	4,91%	A20	270	4,44%
A17	1	0,08%	A12	148	1,99%	A16	157	2,58%
A4	1	0,08%	A13	100	1,34%	A12	119	1,96%
A7	0	0,00%	A20	91	1,22%	A13	95	1,56%
A9	0	0,00%	A14	32	0,43%	A14	14	0,23%
A11	0	0,00%	A9	29	0,39%	A9	2	0,03%

Таблица 1. Статистика распределения FTD по источникам

В Яндекс.Дзен инструктивные тексты (A7) представляют 26.25%, новости (A8) - 18.71%, аргументативные (A1) - 17.1%, справочные (A16) - 11.01%, личные истории (A11) - 9.35%. В текстах ЖЖ ситуация иная: личные истории (A11) - 38.99%, аргументативные (A1) - 20.14%, художественные (A4) - 9.91%. Также особенностью данных источников является тот факт, что очень слабо представлены научные тексты (A14) и нормативно-правовые (A9), а по той причине, что реклама (A12) достаточно часто может быть представлена в форме рецензии (A17) или достаточно схожа с апеллятивными текстами (A20), то ее частотность не так высока. Исходя из этих причин было принято решение расширить материал, добавив в него больше текстов данных типов, чтобы можно было получить более достоверные данные на следующих этапах исследования и рассмотреть такие жанры, как научные и нормативно-правовые тексты более объективно. В итоге в материал были добавлены тексты, которые были отобраны как наиболее близкие к прототипам данных функциональных размерностей, из корпуса ruWAC (Sharoff, 2021): 200 нормативно правовых, 120 научных, 81 рекламных текстов.

Одним из путей доказательства адекватности модели функциональных размерностей для жанровой классификации текстов необходимо выявить особенности текстов, характерных для каждой размерности. Подобные данные позволят дифференцировать FTD, а также подтвердить необходимость учета жанра при анализе текста.

Выбор и получение данных по признакам

На основе программы, разработанной ранее А.Ю. Катинской MDRus Analyser (Катинская, 2016) на Python для идентификации 63 лингвистических признаков, в основном опираясь на (Biber, 1988), и их подсчета в текстах рассматриваемого корпуса, была разработана программа для выявления 71 признака несколько по иному принципу. Мы расширили общий список признаков и несколько доработали некоторые элементы программы (дополнили словари и отредактировали большую часть кода). Так как MDRus Analyser принимает на вход морфологически размеченный через RFTagger корпус, то было принято решение расширить возможности программы и улучшить точность извлечения признаков путем использования другой программы для разметки. Для этого было необходимо рассматривать также синтаксическую разметку для дальнейшего анализа, поэтому программа принимает на вход тексты, размеченные UDPipe, значит, при выявлении отдельных признаков можно опираться не только на морфологические характеристики, частеречную разметку и порядок слов, но и на синтаксические связи, роли и зависимости, что позволяет почти полностью отказаться от привязанности к порядку слов. UDPipe выполняет сегментацию предложений, токенизацию, тегирование по частям речи, лемматизацию, морфологический и синтаксический анализ зависимостей.

Дифференцирование FTD по признакам

Личные истории (A11) - Художественные тексты (A4)

Выявленные значительные пересечения признаков A11 и A4 исходит из того факта, что в обоих случаях зачастую идет повествование, а нарратив с описательными элементами

отражается в значительной частотности местоимений, глаголов прошедшего времени, наречий места, сравнений и т.д. В то же время, A11 и A4 могут показывать признаки A1, по причине того, что подобные виды нарратива также могут подразумевать подведение к выражению точки зрения на отдельную проблему. Из данного факта исходит частотность повелительного и условного наклонения, придаточных со значением причины, и т.д. (см. Таблица 4 в Приложении).

Так как в A4 история личная, то зачастую присутствуют различные средства экспрессивности такие как, например придаточные уступки.

Нарратив предполагает более частотное использование глаголов, а значит, меньше существительных, в особенности неодушевленных. В художественных текстах присутствуют элементы описания, по этой причине нет такого смещения относительно существительных. В то время как в художественных текстах средняя длина предложения сильно зависит от типа текста, а также стиля автора, то значения по статистике средние. Личные же истории ничем не ограничены и не редактируются, а некоторые обычные пользователи интернета имеют тенденцию пренебрежения знаками препинания.

Реклама (A12) - Апеллятивные тексты (A20):

Как в рекламе, так и в апеллятивных текстах основная задача побудить читателя купить что-либо\принять участие в чем-либо. Так как в апеллятивных текстах подразумевается будущее участие в мероприятии, а в результате приобретения товара\услуги должно произойти изменение к лучшему, то статистика по глаголам изменения противоположна (см. Таблица 5 в Приложении).

Описание товара\услуги предполагает оценочную лексику, прилагательные различных типов, неодушевленные существительные. Для привлечения внимания читателя используются местоимения 2 лица и модальные глаголы со значением возможности. В случае же апеллятивных текстов привлечение внимания производится с помощью восклицательных предложений, а так как данные тексты так сильно не привязаны к шаблонам, то длина предложения не ограничивается.

Инструктивные тексты (A7) - Рецензии (A17):

Так как в A7 в основном входят советы, то присутствует непосредственное обращение к читателю (с использованием местоимений 2 лица и возвратных местоимений), инструкции зачастую подаются в форме списков (в этом плане для структуризации помогают вводные конструкции акцентуации) коротких описаний действий (частотно применяется совершенный вид, инфинитив, императив, настоящее время, глаголы изменения, модальные возможности и необходимости). В отдельных случаях присутствует описание проблематики, что отражается в частотности придаточных условия, причины, а также каузативных глаголы. Отсутствие описания, лаконичность и упрощенность ведет к не частотности эффекта крысолова, прилагательных, низкому TTR (см. Таблица 5 в Приложении).

По причине того, что отзывы стилистически не ограничены, то для данной FTD не наблюдается значительного количества выделяющихся признаков. Однако присутствует частотность местоимений, что может быть объяснено как наличием повествования: пересказ части сюжета, а также возможен отрывок личной истории. Так как в отзыве обязательно присутствует оценка какого-либо объекта, то наблюдается значительная частотность оценочной лексики, вводных конструкций со значением оценки, придаточных со значением уступки, наречий и изъяснительных придаточных.

Аргументативные тексты (A1), Новости (A8) и Пропаганда (A13):

Аргументативные и новостные тексты представляют собой самую значительную часть интернет текстов, по этой причине они также широко представлены в рассматриваемом материале (2515 аргументативных текстов, 3147 новостных). Пропаганда была достаточно не частотная на рассматриваемом материале (223 текста), однако возможно выявить некоторые особенности.

A13 можно рассматривать скорее, как часть A1, так как в этом случае в данную размерность входят скорее аргументативные тексты на политическую тематику, а потому наблюдается пересечение признаков. То есть описание личной точки зрения предполагает широкое использование придаточных причины, изъяснительных придаточных, условного наклонения и глаголов речи. Так как в пропаганду входит значительное количество текстов, где описывается, что было сказано\сделано в политической сфере и смежных, то это ведет к частотности совершенного вида, глаголов речи и относительных придаточных (см. Таблица 6 в Приложении).

Описание непосредственной позиции производится за счёт использования инфинитива, глаголов существования, возвратных местоимений, настоящего времени, наречий, придаточных со значением причины, изъяснительных придаточных, условного наклонения. Есть также некоторая тенденция к дистанцированию с помощью глаголов речи, ментальные глаголов, эвиденциальности. В текстах данного жанра распространена экспрессивность, которая выражается в частотности вводных конструкций со значением выделения сказанного, а также вопросительных предложений, частиц.

Так как в новостях происходит описание уже прошедшего или будущего события, а также интервью, то для данной цели часто используются вводные конструкции со значением отношения какого-то отрезка текста к общей последовательности мысли. События новостей в той или иной мере касаются непосредственно определенных людей, что ведет к частотности одушевленных существительных, действительных причастных оборотов, адъективных определений.

Нормативно-правовые тексты (A9), Научные тексты (A14) и Справочные тексты (A16)

Данных FTD в рассматриваемом материале не так много исходя из особенностей источников, однако эти жанры сильно выделяются по сравнению с предыдущими.

В A9 присутствуют многие устойчивые для данной тематики структуры, такие как номинализация, страдательный залог и пр. Так как подобные тексты должны покрывать большое количество ситуаций, то есть некоторая абстрактность, выражающаяся с помощью абстрактных существительных, существительных и наречий со значением времени, существительных с количественным значением (см. Таблица 7 в Приложении). Однако данный факт не приводит к частотности местоимений, чтобы предложение невозможно было толковать двусмысленно относительно антецедентов.

Оценка описываемых подходов в пределах A14 производится с помощью оценочной лексики, также частотны ссылки на источники, и как итог скобочные конструкции. Для описания проведения исследования широко используется пассив, деепричастия, существительные с количественным значением, числительные, глаголы изменения, прилагательные физических свойств, со значением времени, размера. Тематика и стилистические особенности предполагают значительно более высокую сложность текстов, что как итог приводит к частотности эффекта крысолова, высокому TTR, значительной средней длине слова.

Так как речь идет об одном и том же объекте в пределах текста A16, то частотны местоимения 3 л, а также редко сочинения фразовых категорий. Описание термина занимает значительную часть в текстах данного жанра, поэтому частотны относительные придаточные, прилагательные в предикативной функции, вводные конструкции со значением отношения отрезка текста к общей последовательности мысли, глаголы существования.

Различия в функционировании в текстах разных FTD

Хеджирование.

Исходя из предыдущего пункта рассмотрение хеджирования также имеет больший смысл в случаях, где высказывается какое-либо личное мнение, то есть A1, A11, A17, так

как в данных случаях авторы имеют тенденцию дистанцироваться от высказывания с различными целями. Совсем не выражено использование данного приема в А8, А9, А12, А14, А16, что ожидаемо в связи с тем, что во всех случаях на первом плане стоит информирование.

	indefinite pron	mental verb	negation	modal possibility	modal necessity	evidentiality	parenthesis attitude evaluation
A1	0.00385	0.00389	0.01654	0.00349	0.00129	0.001232	0.00075
A4	0.00319	0.00296	0.01808	0.00186	0.00073	0.001153	0.00035
A7	0.00223	0.00333	0.01322	0.00487	0.00311	0.000440	0.00046
A8	0.00092	0.00154	0.00689	0.00181	0.00052	0.000494	0.00017
A9	0.00146	0.00369	0.00616	0.00286	0.00089	0.000075	0.00001
A11	0.00418	0.00363	0.01757	0.00296	0.00091	0.001359	0.00077
A12	0.00206	0.00285	0.00653	0.00404	0.00073	0.000276	0.00026
A13	0.00257	0.00263	0.01403	0.00247	0.00107	0.000821	0.00065
A14	0.00128	0.00368	0.00532	0.00314	0.00060	0.000437	0.00015
A16	0.00191	0.00255	0.00952	0.00323	0.00086	0.000605	0.00040
A17	0.00306	0.00311	0.01484	0.00385	0.00095	0.001073	0.00095
A20	0.00220	0.00272	0.00892	0.00317	0.00147	0.000561	0.00077

Таблица 2. Относительные частоты выражения хеджирования

В художественных текстах (А4) не характерно использование модальности, однако частотна эвиденциальность, отрицание и неопределенные местоимения, что отчасти объясняется тематикой, но данные могут различаться в этом плане относительно материала данного жанра.

Средние показатели в целом и высокие по отдельным способам выражения хеджирования показывают размерности инструктивные тексты (А7), пропаганда (А13), апеллятивные тексты (А20). В первом случае основные функции текстов определяют частотность модальности, во втором - эвиденциальности и вводных конструкций со значением оценки, в третьем - модальности необходимости и вводных конструкций со значением оценки.

Количественные группы типа “две серьезные аварии / две серьезных аварии”

В работе [Nesset, 2019] автор строит статистические модели, описывающие устройство русских количественных словосочетаний типа две серьезные/ых аварии. Один из аспектов, которые рассматривались относительно данной конструкции следующий: относительная частота прилагательных в родительном падеже выше для “три” и “четыре”, чем для “две”. Далее рассмотрим, есть ли какие-либо вариации в этом плане относительно жанров на сравнительно более небольшом материале.

Так как материала А9, А12, А13, А14, А20 не так много, то рассматривать данное явление на текстах данных жанров в пределах исследования нет смысла, так как частоты очень низкие. Также исключается размерность А4 по причине того, что в этом случае частотность подобных конструкций также низка, но уже не по причине недостатка в материале, а по причине низкой частотности числительных.

	A1		A7		A8		A11		A16		A17	
	nom	gen	nom	gen	nom	gen	nom	gen	nom	gen	nom	gen
две	75,00%	25,00%	89,36%	10,64%	88,00%	12,00%	80,95%	19,05%	72,73%	27,27%	83,33%	16,67%
	24	8	42	5	22	3	17	4	40	15	10	2

три	10,00%	90,00%	37,31%	62,69%	35,29%	64,71%	19,05%	80,95%	44,44%	55,56%	25,00%	75,00%
	2	18	25	42	12	22	4	17	28	35	2	6
четыре	14,29%	85,71%	14,71%	85,29%	9,09%	90,91%	0,00%	100,00%	44,74%	55,26%	0,00%	100,00%
	1	6	5	29	2	20	0	2	17	21	0	4

Таблица 3. Абсолютные частоты вариантов *nom\gen*

В целом данные схожие для всех жанров и серьезно не противоречат предыдущим работам относительно данного вопроса, но есть ряд особенностей:

- 1) Предпочтение номинатива при “две” максимальное в инструктивных (A7) и новостных текстах (A8), почти 90%.
- 2) В FTD A1, A11, A17 в случаях с “три” основная тенденция наиболее ярко выражена, чем в других (A1, A11, A17 – 75%-90% генетива к около 60% в A7, A8).
- 3) В справочной информации (A16) в случае в “три” частоты падежей примерно равны.

Заключение

На примере данных базовых характеристик текстов можно наблюдать разброс в частотности до нескольких десятков раз в отдельных признаках даже в достаточно близких по природе размерностях. Так как даже простые признаки могут так значительно различаться, то статистика каких-либо иных более сложных конструкций и явлений будет также показывать перекосы относительно размерностей. По этой причине получить наиболее достоверные результаты при проведении исследования можно либо на основе жанрово однородного материала, либо требуется сравнивать разные жанрово однородные сегменты между собой.

Литература

1. Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge University Press.
2. Biber, D. (1990) Methodological issues regarding corpus-based analyses of linguistic variation. // *Literary and Linguistic computing*, vol. 5, no. 4, Oxford University Press.
3. Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press.
4. Biber, D. and Egbert, J. (2016). Register variation on the searchable web: A multi-dimensional analysis. *Journal of English Linguistics*, 44(2):95–137.
5. Biber, D. and Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
6. Katinskaya, A. and Sharoff, S. (2016). Applying multi-dimensional analysis to a Russian webcorpus: Searching for evidence of genres. In *Proc BSNLP*, Sofia.
7. Lagutin (2015) M. B., Katinskaya A. Y., Selegey V. P., Sharoff S., Sorokin A. A. Automatic classification of web texts using functional text dimensions
8. Nessel, T. (2019), Big data in Russian linguistics? *Zeitschrift Für Slawistik*, vol. 64(2), pp. 157–174.
9. Sharoff, S (2018) Functional Text Dimensions for the Annotation of Web Corpora. *Corpora*. 13. 65-95. 10.3366/cor.2018.0136.
10. Sharoff, S (2021) Genre Annotation for the Web: Text-external and text-internal perspectives. *Register studies*, Volume 3

Приложение

A11	A4
+first_person_pronoun (0.0199) _(0.0027) _(0.0101) +nom_pronoun (0.0020) +indefinite_pron (0.0041) +past_tense (0.0353) +place_adverb (0.0065) +time_adverb(0.0163) +total_adverb(0.1233) _(0.0032) -nominalization(0.0082) -nouns(0.2051) _(0.0007) +that_compl(0.0084) _(0.0003) _(0.0057) -word_length(4.2042) _(15.0669) +sentence_length(0.7504) _(0.0007) -passive_participial_clauses(0.0011) +active_participial_clauses(0.0243) +imperative_mood(0.0243) -predicative_adjectives(0.0038) -attributive_adjective(0.0316) +causative_subordinate(0.0011) +concessive_subordinate(0.0008) _(0.0002) +negation(0.0175) _(0.0012) _(0.0029) +evidentiality(0.0013) -animate_nouns(0.0296) -parenthesis_relation(0.00007) +other_coordination(0.0322) +degree_adverb(0.0047) +particles(0.0268) +time_nouns(0.0122) _(0.0048) _(0.0174) +movement_verb(0.0081) _(0.0050) +q_particles(0.00093) +unfinished_clauses(0.0038) _(0.0033) +comparison(0.0070) _(0.0316)	+first_person_pronoun (0.0137) +second_person_pronoun (0.0073) +third_person_pronoun(0.0156) _(0.0015) _(0.0031) +past_tense(0.0363) +place_adverb (0.0057) _(0.0107) _(0.1089) +wh_questions(0.0056) -nominalization(0.0079) _(0.2205) -by_passive(0.0006) _(0.0065) -pied_piping(0.0002) +exclamation(0.0843) -word_length (4.1538) -type_token_ratio(15.0669) _(0.7144) +verbal_adverb(0.0011) -passive_participial_clauses(0.0011) +active_participial_clauses(0.0279) +imperative_mood(0.0092) -predicative_adjectives(0.0039) -attributive_adjective(0.0299) _(0.0007) _(0.0005) +purpose_subordinate(0.0004) +negation(0.0180) +conditional_mood(0.0013) -modal_possibility(0.0018) +evidentiality(0.0011) _(0.0405) -parenthesis_relation(0.00004) +other_coordination(0.0338) _(0.0026) +particles(0.0256) _(0.0087) -quantity_nouns(0.0048) -numeral(0.0096) +movement_verb(0.0072) +parceling(0.0073) +q_particles(0.00096) +unfinished_clauses(0.0035) -parentheses(0.0017) +comparison(0.0033) +verbless_clauses(0.0370)

Таблица 4. Признаки A11, A4

A12	A20	A7	A17
- third_person_pronoun(0.0050) -past_tense(0.0128) +phrasal_coordination(0.0592) -time_nouns(0.0071)	- third_person_pronoun(0.0050) -past_tense(0.0125) +phrasal_coordination(0.0592) -time_nouns(0.0105)	+second_person_pronoun(0.0065) +reflexive_pronoun(0.0022)) -past_tense(0.0058) +perf_aspect(0.0494) +present_tense(0.0445)	+first_person_pronoun(0.0090) +third_person_pronoun(0.0096) +adjective_pronoun(0.0058)) +total_adverb(0.1176)

+change_verb(0.0014) +second_person_pronoun(0.0055) -speech_verb(0.0012) -that_compl(0.0025) - active_participial_clauses(0.0076) +predicative_adjectives(0.0052) +modal_possibility(0.0040) +evaluative_vocabulary(0.00877) -animate_nouns(0.0231) -movement_verb(0.0045) +physical_prop_adjective(0.0044) +size_adjective(0.0044)	-change_verb(0.0004) -passive(0.0046) +exclamation(0.0082) +sentence_length(0.7825) -verbal_adverb(0.0004) - parenthesis_accentuation(0.0001)	+infin(0.1176) -pied_piping(0.0002) - -type_token_ratio(15.432) +imperative_mood(0.0083) - predicative_adjectives(0.0039) +conditional_subordinate(0.0039) +purpose_subordinate(0.0005) +modal_possibility(0.0048) +modal_necessity(0.0031) -animate_nouns(0.0192) +parenthesis_accentuation(0.0007) +causative_verb(0.0139) +change_verb(0.0020) -time_adjective(0.0017) +parceling(0.0066)	+wh_relative(0.0021) +concessive_subordinate(0.0009) +parenthesis_attitude_evaluation(0.0009) +evaluative_vocabulary(0.00791) +degree_adverb(0.0057)
--	--	--	--

Таблица 5. Признаки A12, A20, A7, A17

A1	A13	A8
+that_compl(0.0086) +causative_subordinate (0.0012) +conditional_mood (0.0014) +speech_verb(0.0028) +evidentiality(0.0012) -quantity_nouns(0.0055) +q_particles(0.0011) +comparison(0.0072) +reflexive_pronoun (0.0025) +adjective_pronoun (0.0060) +present_tense(0.0417) +total_adverb(0.1112) +infin(0.0246) +mental_verb(0.0038) +parenthesis_accentuation(0.0008) +existential_verb(0.0066) +nom_pronoun(0.0060) +indefinite_pro(0.0038) +wh_questions(0.0050)	+that_compl(0.0102) +causative_subordinate (0.0011) +conditional_mood(0.0015) +speech_verb(0.0027) +animate_nouns(0.0562) +wh_relative(0.0022) -phrasal_coordination(0.0451) -quantity_nouns(0.0057) -phisical_prop_adjective (0.0010)	-second_person_pronoun (0.0005) -reflexive_pronoun(0.0009) -adjective_pronoun(0.0026) -nom_pronoun (0.0008) -indefinite_pron(0.0009) -present_tense(0.0279) -wh_questions(0.0006) +nouns (0.3254) -passive(0.0044) -by_passive(0.0006) -mental_verb(0.0015) -pied_piping(0.0002) -total_PP(0.1026) -exclamation(0.0007) -verbal_adverb(0.0003) +active_participial_clauses (0.0237) +predicative_adjectives (0.0050) -conditional_subordinate (0.0007) -modal_possibility(0.0018) -modal_necessity(0.0005) -evaluative_vocabulary (0.0339) +animate_nouns(0.0535) -parenthesis_accentuation (0.0002) +parenthesis_relation(0.0004) -particles(0.0090) -causative_verb(0.0054) -existential_verb(0.0033) -movement_verb(0.0042) -phisical_prop_adjective (0.0009) -parataxis(0.0033) -abstract_nouns(0.0629)

Таблица 6. Признаки A1, A13, A8

A9	A14	A16
-first_person_pronoun (0.0006) -third_person_pronoun (0.0060)	-first_person_pronoun (0.0015) -second_person_pronoun (0.0004)	+third_person_pronoun (0.0125) -nouns(0.2962)

<p>-adjective_pronoun(0.0025) -nom_pronoun (0.0008) -perf_aspect(0.0274) -place_adverb(0.0005) -time_adverb(0.0019) -total_adverb(0.0176) -wh_questions(0.0003) +nominalization(0.0541) +passive(0.0265) +by_passive(0.0046) -infin(0.0118) -that_compl(0.0015) -wh_relative(0.0012) +pied_piping(0.0010) -total_PP(0.1044) -exclamation(0.0003) +word_length(5.7783) +type_token_ratio(41.456) -sentence_length(0.4942) +passive_participial_clauses (0.0180) -imperative_mood(0.0004) -predicative_adjectives (0.0040) -attributive_adjective(0.1062) -causative_subordinate (0.0003) -concessive_subordinate (0.00004) -purpose_subordinate (0.00007) -negation(0.0062) -conditional_mood(0.00005) -evidentiality(0.00008) - parenthesis_attitude_evaluatio(0.0001) - parenthesis_accentuation (0.0002) -other_coordination(0.0063) -degree_adverb(0.0002) -particles(0.0095) +time_nouns(0.0176) +quantity_nouns(0.0146) +numeral(0.0531) -change_verb(0.0004) +time_adjective(0.0045) -size_adjective (0.0008) -parceling(0.0020) -q_particles(0.0001) +parentheses(0.0102) +parataxis(0.0158) +abstract_nouns(0.2190) -comparison(0.0014) -verbless_clauses(0.0137)</p>	<p>-indefinite_pron(0.0012) -perf_aspect(0.0294) -time_adverb(0.0040) -total_adverb(0.0382) -wh_questions(0.0005) +passive(0.0222) +by_passive(0.0035) -infin(0.0109) +pied_piping(0.0008) -exclamation(0.0004) +word_length(5.5070) +type_token_ratio(24.522) +verbal_adverb(0.0011) +passive_participial_clauses (0.0126) -imperative_mood(0.0009) -attributive_adjective(0.1006) -conditional_subordinate (0.0008) -negation(0.0053) -modal_necessity(0.0006) +evaluative_vocabulary (0.0900) +parenthesis_relation(0.0004) -other_coordination(0.0081) +quantity_nouns(0.0165) +numeral(0.0458) +change_verb(0.0015) +phisical_prop_adjective (0.0049) +time_adjective(0.0046) +size_adjective(0.0040) -parceling(0.0016) -q_particles(0.0001) -unfinished_clauses(0.00007) +parentheses(0.0120) +parataxis(0.0183) +abstract_nouns(0.1688) -verbless_clauses(0.0174)</p>	<p>+wh_relative(0.0025) +predicative_adjectives (0.0056) +parenthesis_relation(0.0004) +existential_verb(0.0003) -phrasal_coordination(0.0458)</p>
--	--	--

Таблица 7. Признаки A9, A14, A16