

УДК 81-25

«КАК БЫ ОН ЭКСТРАВЕРТ ЧТО ЛИ?»: К ВОПРОСУ О КОРРЕЛЯЦИИ ПРАГМАТИЧЕСКИХ МАРКЕРОВ-АППРОКСИМАТОРОВ С ПСИХОТИПОМ ЧЕЛОВЕКА¹

Дарья Александровна Горбунова

аспирант кафедры русского языка

Санкт-Петербургский государственный университет

199034, г. Санкт-Петербург, Университетская наб., 11. st058398@student.spbu.ru

Статья посвящена исследованию связей между речевыми единицами и психологическим типом человека (экстраверт/интроверт). Спонтанные вербальные высказывания были проанализированы на предмет наличия в них маркеров нечеткой номинации (аппроксиматоров *как бы, вроде, своего рода, типа, там* и т. д.). В рамках исследования был использован набор методов Data Mining для подробного статистического анализа корреляции употреблений прагматических маркеров выделенного типа с психотипом говорящего. Данные (3 661 единица в контекстах, 300 000 токенов) были собраны на основании всех размеченных прагматических маркеров и метаданных каждого информанта из выборки. Результаты анализа этих данных показывают значительную положительную корреляцию между психологическим типом говорящего и маркерами нечеткой номинации, которые он использует в своей повседневной речи.

Ключевые слова: прагматический маркер; психолингвистика; коллоквиалистика; психотип; устная речь; компьютерная лингвистика.

Введение

В отечественной и зарубежной психолингвистике уже довольно давно ведется разработка такой проблемы, как установление связи особенностей речи говорящих с их психологическими характеристиками, однако количество исследований в этой области ограничено. Как правило, исследователей больше интересует речь как когнитивный процесс, а также процессы, протекающие совместно с речью и зависящие от нее, такие как обучение, память, восприятие речи. Речь с этой точки зрения неизменно рассматривается как простое следствие биологических и когнитивных процессов [Furnham 1990: 74]. Второй причиной отсутствия большого количества исследований корреляций между личностью говорящего и его речью, по мнению А. Фернхэма, является сложность в поиске подходящего уровня для анализа личности с языковой точки зрения [там же: 75]. В своей более поздней работе, написанной в соавторстве с Ж.-М. Деваэле, А. Фернхэм добавляет, что «многие исследователи психологических типов личности совершенно не знакомы с прикладной лингвистикой или не обладают достаточной квалификацией, чтобы выбирать и затем измерять незначительные, но все же важные языковые индикаторы в разговор-

ной речи, отражающие, например, беглость, формальность, сложность, лексическое богатство, морфологическую точность и социостилистическую компетентность»² [Dewaele, Furnham 1999: 519].

Дж. Пенненбейкер и Л.А. Кинг исследовали личностные черты в языковом контексте и выделили категории слов, наиболее соответствующие экстраверсии: эмоционально-экспрессивную лексику, личные местоимения первого лица и глаголы настоящего времени [Pennebaker, King 1999]. Однако исследование проводилось на письменном материале, и речевая реальность, скорее всего, внесла бы коррективы в данные о частотности этих категорий. Ценность этого исследования заключается прежде всего в том, что вопреки мнению А. Фернхэма, был найден уровень анализа личности с языковой точки зрения: наиболее целесообразным представляется изучение личностных характеристик говорящего на базе языковых единиц, которые отличаются крайней распространенностью в речи всех говорящих.

Такими единицами в настоящем исследовании стали функциональные единицы русской устной речи, которые мы называем **прагматическими маркерами** (ПМ) и не ассоциируем с

классом **дискурсивных маркеров** (ДМ), достаточно хорошо описанных как в отечественной, так и в зарубежной литературе (см., например, обзор соответствующих работ в [Зайдес 2019; Bogdanova-Beglarian, Filyasova 2018; Bogdanova-Beglarian et al. 2018a, 2018b]). Прагматические маркеры, как правило, практически лишены как лексического, так и грамматического значения, а значит, их употребление можно считать если не всецело обусловленным психологически, то, предположительно, имеющим корреляцию с психологическим типом говорящего.

Методы и материал исследования

Источником лингвистического и психологического материала для анализа послужил корпус повседневной русской речи «Один речевой день» (ОРД) – один из наиболее представительных современных ресурсов для исследования русского устного дискурса [Русский язык повседневного общения 2016; Asinovsky et al. 2009; Bogdanova-Beglarian et al. 2019].

Анализ размеченных корпусных данных проводился при помощи набора методов Data Mining (см. подробнее: [Степанов 2008]), которые при помощи инструментов математической статистики производить поиск зависимостей между всеми типами представленных данных.

Психологический тип информантов ОРД был определен при помощи специально созданного программного инструментария, сопоставляющего результаты двух психологических тестов – EPI и FPI (см. подробнее об этом: [Gorbunova 2017]) – по всем используемым в них шкалам. Оба психологических теста заполнили только 68 человек из более чем ста информантов корпуса. Совпадение результатов тестов по шкале интроверсия/экстраверсия составило 62%. Датасет (набор данных для анализа) в настоящем исследовании был собран из расшифровок «речевых дней» только этих информантов. Для анализа датасета (3 661 единица в контекстах, 300 000 токенов) был использован язык программирования Python 3.6 (библиотеки Pandas и NumPy). Данные прошли несколько стадий предварительной обработки:

- 1) перевод всех букв в расшифровках в нижний регистр;
- 2) удаление всех знаков расшифровки;
- 3) автоматическое распознавание и удаление слов коммуникантов;
- 4) автоматический подсчет количества слов информантов;
- 5) «склеивание» кода информанта, его психотипа и всех его реплик;
- 6) фильтрация реплик, содержащих прагматический маркер;

7) ручная разметка функционального типа ПМ в репликах.

Поскольку при разметке неизменно обнаруживались полифункциональные ПМ, выполняющие сразу две или три функции в одном и том же контексте, было принято решение при подсчетах разделить такие употребления и рассматривать один и тот же ПМ в разных функциях как две или более единицы. Например:

- (1) *я / (...) как бы / я не знаю / у них какое-то ... / ну просто не досмотрели чего-то ребята* (И101, ЭТ)³.

Прагматический маркер *как бы* из контекста (1) проаннотированный как АХ (аппроксиматор/хезитатив), рассматривался как два употребления ПМ: *как бы* (X) и *как бы* (A).

В результате предварительной обработки был получен финальный набор данных, состоящий из 109 677 слов и 1 499 контекстов с ПМ различных функциональных типов.

Основной задачей дальнейшего анализа стал поиск ассоциативных правил [Степанов 2008: 10] – частных зависимостей (или ассоциаций) между объектами или событиями, в данном случае между маркерами-аппроксиматорами и психологическим типом информанта.

Маркеры-аппроксиматоры и их корреляция с психотипом человека

Аппроксиматоры являются маркерами нечеткой, или приблизительной, номинации, которые употребляются говорящим, когда прямое название предмета, явления или положения дел является излишним, неуместным или невозможным:

- (2) *значит (...) как бы нырять / и плыть вперед* (И120, ЭТ);
- (3) *то есть было оборудование было тиски были как бы ну инструментари* (И102, ЭТ).

Однако остается неясным, являются ли маркеры-аппроксиматоры определяющими при соотношении психологических типов и можно ли по их наличию или отсутствию в речи определить индивидов с относительно высоким или низким уровнем экстраверсии. Первое, на что важно обратить внимание, это частотность употребления маркеров. Ниже представлены результаты автоматического анализа частотности маркеров-аппроксиматоров и показано их распределение по психологическим типам информантов – интроверт/экстраверт (см. табл.), где

- p-type – психотип говорящего (знак «→» маркирует информантов, не заполнивших один или несколько тестов);
- Function – функция ПМ;
- NContexts – количество контекстов, в которых ПМ употребляется в указанной функции;

– %TotalPtypeWords – отношение количества ПМ в указанной функции к общему количеству слов⁴ в речи говорящих с данным психотипом;

– %TotalPtypeContexts – отношение количества контекстов с ПМ в указанной функции к общему количеству контекстов в речи говорящих с данным психотипом.

Таблица

Распределение употреблений маркеров-аппроксиматоров по психотипу говорящих

Function	p-type	NContexts	%TotalPtypeWords	%TotalPtypeContext
А	–	152	0,09	7,6
	интроверт	22	0,07	3,9
	экстраверт	62	0,07	6,5

При проведении многоаспектного функционального анализа важно ориентироваться на отношение количества ПМ в указанной функции к общему количеству слов в речи говорящих с данным психотипом, поскольку интроверты и экстраверты использовали в течение своих «речевых дней» разное по объему количество слов. В этом случае количество употреблений маркеров-аппроксиматоров у интровертов и экстравертов совпадает – по 0,07%. Но поскольку большая часть маркеров-аппроксиматоров неоднословны, статистически целесообразнее опираться на отношение количества контекстов с ПМ в указанной функции к общему количеству контекстов в речи говорящих с данным психотипом. И здесь можно видеть значительную разницу в количестве употреблений данного типа ПМ: 6,5% у экстравертов; 3,9% у интровертов, что практически в половину меньше, чем у экстравертов. Проанализируем несколько частных случаев таких употреблений:

(4) *то есть это уже как бы для нас не важно / в какую сторону кровь идет* (И130, ИТ);

(5) *так что он как бы потомственный лесо (...) вод и вот (м-м) () любит он дерево / любит дерево // и вообще деревянная тема у него преобладает* (И147, ИТ).

Заметно, что интроверты в контекстах (4) и (5) используют *как бы* в качестве своеобразной речевой «разминки», поначалу нечетко определяя предмет разговора с помощью маркера-аппроксиматора, нащупывая дальнейший путь и с каждым словом все более уточняя и поясняя то, о чем идет речь. Не уверенный в своих речевых навыках, собеседник-интроверт стремится как можно более полно объяснить то, что он имеет в виду, и не допустить вольной трактовки случайно брошенного *как бы*.

(6) *гитарист нужен в любой репертуаре вообще / как бы вот осваивать* (И125, ЭТ);

(7) *ну:(.) () там () типа / каталог заранее сделан / вроде бы конкретные работы уже приедут* (И121, ЭТ);

Экстраверты в контекстах (6) и (7) при помощи маркеров-аппроксиматоров добавляют не-

определенности, выражают некоторое сомнение, стремятся к абстрактному отображению реальности. Дж. Ченнелл предположила, что существуют различные средства, с помощью которых говорящие передают неопределенность: импликацией, выбором расплывчатых слов. Исследователь утверждает, что неявная неясность возникает из представления о том, что даже явно точные фразы могут быть истолкованы как имеющие неопределенное значение. С другой стороны, «нечеткие добавки» (маркеры-аппроксиматоры) «представляют собой лексические элементы, которые изменяют степень определенности или прямолинейности составляющих, которые они модифицируют»⁵ [Channell 1994: 36]. В следующем контексте (8) реализована как раз такая, характерная для экстравертов смена степени определенности при помощи маркера-аппроксиматора *вроде как*:

(8) *нет (э-э) г... () потом тоже спрашивали / вроде как нет / но тем не менее / дождь (э-э) где-то в другой части города был *П всю ночь / а на Дворцовой\$-то не было* (И65, ЭТ).

Аппроксиматор часто сопровождает передачу чужой речи или просто информации, полученной от другого человека, т. е. выступает в функции *ксенопоказателя*, или ввода *ксенопоказателя*. В этом случае информация потенциально недостоверна именно по этой причине. Таких случаев гораздо больше в речи экстравертов. Аппроксиматор может сопровождать глагол говорения, как в примере (8), при этом маркер может присоединяться как к глаголу (говорящий сигнализирует о своих сомнениях в самом факте сообщения), так и к ренарративу, как в примере (9):

(9) *ну:(.) вообще мы как (...) обговаривали / что вроде бы да // *П а я просто позволю тогда Татьяна% и... и уточню* (И113, ЭТ).

Заключение

Исследование показало, что экстраверсия как психологический тип личности говорящего коррелирует с употреблением маркеров-аппроксиматоров в устной речи. Это дает дополнительные

доказательства того, что экстраверты склонны говорить более абстрактно, меняя степень определенности своих высказываний при помощи маркеров-аппроксиматоров. Интроверты, в свою очередь, используют такие маркеры как основу для развернутых описательных конструкций. С точки зрения психологии это вполне объяснимо, поскольку интроверты обычно высказывают свое мнение осторожно, используя маркеры нечеткой номинации, однако стремятся передать мысль наиболее точно. Экстраверты в речи самоуверенны, однако выражают свои мысли более спонтанно, часто переходя от одной степени определенности к другой.

В перспективе настоящего исследования – расширение датасета и выявление более специфических лингвистических признаков для статистического прогнозирования (определения объективной вероятности) психологического типа на основании употребления маркеров-аппроксиматоров. Датасет и результаты текущего исследования могут быть использованы при создании алгоритмов обработки естественного языка (Natural Language Processing), а также систем автоматического синтеза и распознавания речи.

Примечания

¹ Исследование выполнено при поддержке гранта РФФИ № 18-18-00242 «Система прагматических маркеров русской повседневной речи»

² «Of course, it must also be pointed out that many personality researchers are quite ignorant of, or not qualified in, applied linguistics to choose and then measure subtle, yet important, language indicators in spoken language reflecting, for example, fluency, formality, complexity, lexical richness, morphological accuracy, and sociostylistic competence» [Dewaele, Furnham 1999: 519] (здесь и далее перевод мой. – Д. Г.).

³ Об особенностях представления материала в корпусе ОРД см.: [Шерстинова Рыко, Степанова 2009; Русский язык повседневного общения 2016: 242–243]. Все примеры из этого корпуса в статье атрибутируются с указанием номера информанта (например И120) и его психотипа, где ЭТ – экстраверт, а ИТ – интроверт.

⁴ Под количеством слов понимается результат автоматического подсчета, при котором учитывались только слова информанта, без слов его коммуниканта(ов), помет и знаков расшифровки.

⁵ «...represent lexical items that alter the degree of certainty or straightforwardness of the constituents they modify» [Channell 1994: 36].

Список литературы

Зайдес К.Д. Прагматические маркеры предикативного типа в устной спонтанной речи: подходы к описанию // Коммуникативные исследования. 2019. № 2(16). С. 375–396.

Русский язык повседневного общения: особенности функционирования в разных социальных группах / отв. ред. Н.В. Богданова-Бегларян. СПб.: ЛАЙКА, 2016. 244 с.

Степанов Р.Г. Технология Data Mining: интеллектуальный анализ данных. Казань: Казанский гос. ун-т им. В. И. Ульянова-Ленина, 2008. 58 с.

Шерстинова Т.Ю., Рыко А.И., Степанова С.Б. Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Формальные методы анализа речи: материалы XXXVIII Междунар. филол. конф. СПб.: Факультет филологии и искусств С.-Петерб.ГУ, 2009. С. 66-75.

Asinovsky A. et al. The ORD Speech Corpus of Russian Everyday Communication «One Speaker's Day»: Creation Principles and Annotation / Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. // Lecture Notes in Artificial Intelligence (LNAI). 2009. V. 5729: Text, Speech and Dialogue: 12th International Conference – TSD 2009 / V. Matoušek, P. Mautner (eds.). P. 250–257.

Bogdanova-Beglarian N. et al. Pragmatic Markers in Russian Spoken Speech: an Experience of Systematization and Annotation for the Improvement of NLP Tasks / Bogdanova-Beglarian N., Blinova O., Sherstinova T., Martynenko G., Zaides K. // Proceedings of the 23rd Conference of Open Innovations Association – FRUCT'23. Bologna, 2018a. P. 69-78.

Bogdanova-Beglarian N. et al. Towards a Description of Pragmatic Markers in Russian Everyday Speech / Bogdanova-Beglarian N., Baeva E., Blinova O., Martynenko G., Sherstinova T. // Lecture Notes in Computer Science (LNCS). 2018b. V. 11096: International Conference on Speech and Computer – SPECOM 2018 / A. Karpov, O. Jokisch, R. Potapova (eds.). P. 42–48.

Bogdanova-Beglarian N. V. et al. Pragmatic Markers of Russian Everyday Speech: the Revised Typology and Corpus-Based Study / Bogdanova-Beglarian N.V., Blinova O.V., Sherstinova T.Ju., Troshchenkova E.V., Gorbunov D.A., Zaides K.D. // Proceedings of the 25th Conference of Open Innovations Association – FRUCT'25 / S. Balandin, V. Niemi, T. Tuytina (eds.). Helsinki, 2019. P. 57–63.

Bogdanova-Beglarian N.V., Filyasova Yu.A. Discourse vs. Pragmatic Markers: A Contrastive Terminological Study // 5th International Multidisciplinary Scientific Conference on Social Sciences and Arts –

SGEM 2018: Proceedings. Vienna, 2018. V. 5, iss. 3(1). P. 123–130.

Channell J. Vague Language (Describing English Language Series). Oxford: Oxford University Press, 1994. 226 p.

Dewaele J.-M., Furnham A. Extraversion: The Unloved Variable in Applied Linguistic Research // Language Learning. 1999. № 19. P. 509–544.

Furnham A. Language and Personality // Handbook of Language and Social Psychology / H. Giles,

W.P. Robinson (eds.). Chichester, UK: John Wiley & Sons, 1990. P. 73–95.

Gorbunova D. Spontaneous Oral Speech From the Psycholinguistic Side: Pragmatemes in the Speech of Informants with a Different Psychotype // Communicative Studies. 2017. № 4(14), P. 73–82.

Pennebaker J.W., King L.A. Linguistic Styles: Language Use as an Individual Difference // Journal of Personality and Social Psychology. 1999. № 77. P. 1296-1312.

“IT’S LIKE HE IS AN EXTRAVERT OR WHAT?”: ON CORRELATION OF PRAGMATIC MARKERS-APPROXIMATORS WITH A PSYCHOTYPE

Darya A. Gorbunova

PhD Student, Russian Language Department
St. Petersburg State University

The article is devoted to the study of direct and indirect connections between speech units and a person’s psychological type (extrovert/introvert). Spontaneous verbal utterances were analyzed for the presence of fuzzy markers (approximators *sort of, kind of, like, there*, etc.). Data Mining methods were used for a detailed statistical analysis of how the use of the pragmatic markers correlates with the speaker’s psychotype. The data (3 661 units in contexts, 300,000 tokens) were collected using all the tagged pragmatic markers and metadata for each informant of the sample. The results of the data analysis show a significant positive correlation between the speaker’s psychological type and the fuzzy markers used in one’s everyday speech.

Keywords: pragmatic marker; psycholinguistics; colloquialistics; psychological type; oral speech; computational linguistics.