

Репрезентации уровня развития личности в тексте: синтаксические и лексические индикаторы

© 2020 Костенко Василий Юрьевич*, Кузьмина Алиса Андреевна**,

*кандидат психологических наук, доцент департамента психологии, старший научный сотрудник Международной лаборатории позитивной психологии личности и мотивации, Национальный исследовательский университет «Высшая школа экономики» (г. Москва), e-mail: vkostenko@hse.ru

** студентка магистерской программы «Позитивная психология», Национальный исследовательский университет «Высшая школа экономики» (г. Москва), e-mail: aakuzmina_2@edu.hse.ru

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 20-78-00107.

Аннотация: Развитие личности на современном этапе его понимания может быть рассмотрено как процесс постепенного усложнения форм и способов интерпретации личностью себя и действительности [Loevinger, 1976]. Операционализацией такого подхода выступает Методика неоконченных предложений Вашингтонского университета (НПВУ). Вербальный материал ответов данного инструмента позволяет проводить его анализ в междисциплинарном поле психологии личности и компьютерной лингвистики. В данном исследовании при помощи стратегии подсчета слов и метода закрытого словаря LIWC анализируется лингвистический корпус из 37 458 завершений предложений. Показано, что проявления в ответах респондентов запятых и кавычек статистически значимо возрастает от одного уровня развития личности к другому. Ключевые лексические индикаторы (общее число

слов в протоколе, среднее число слов в предложении и доля слов длиннее 6 букв) также увеличиваются по мере повышения уровня личностной зрелости. Результаты свидетельствуют в пользу подтверждения общей теоретической гипотезы о проявлении уровней развития личности в способах использования слов.

Ключевые слова: развитие личности, уровень развития эго; LIWC; метод подсчета слов.

**Manifestations of personality development in textual data:
syntactic and lexical indicators**

© 2020 Kostenko Vasily Yuryevich*, Kuzmina Alisa Andreyevna**,

*Ph.D. in Psychology, Associate Professor in School of Psychology, Senior Research Fellow of the International Laboratory of Positive Psychology of Personality and Motivation, National Research University Higher School of Economics (Moscow, Russia), e-mail: vkostenko@hse.ru

**Student of Master's Program of Positive Psychology, National Research University Higher School of Economics (Moscow, Russia),
e-mail: aakuzmina_2@edu.hse.ru

Annotation: Personality development is considered as a process of gradual complication of the forms of interpretation of psychosocial reality. J. Loevinger (1976) developed the Washington University Sentence Completion Test (WUSCT) to measure this dimension. The verbal material of the responses allows analysis within the interdisciplinary field of personality psychology and computational linguistics. In this study, a linguistic corpus of 37,458 sentence completions is analyzed using the word counting strategy and the Linguistic Inquiry and Word Count (LIWC) method. It is shown that the manifestation of commas and quotation marks in the respondents' answers increased significantly from one level of personality development to another. Key lexical indicators (the total number of words in the protocol, the average number

of words in a sentence, and the proportion of words longer than six letters) also increased on higher personality maturity levels. The results confirmed the general hypothesis on the manifestation of personality development levels within the theoretical framework of word usage.

Keywords: personality development, ego development; LIWC; word counting method.

Междисциплинарное поле лингвистики и психологии личности активно развивается благодаря обширному применению методов компьютерной лингвистики и машинного обучения. Изучение количественных морфологических индикаторов психологических феноменов на материале текстовых данных позволило бы уточнить теоретическое понимание некоторых психологических конструктов и расширить возможности их диагностики. Подобная работа в настоящее время проводится в отношении исследования различных психических расстройств [Lyons, Aksayli, Brewer, 2018], феномена переживания приближения смерти [Goranson et al., 2017], терроризма [Vergani, Vliuc, 2018], музыкальных предпочтений [Qiu et al., 2019] и во многих других областях исследований.

Обратившись к методам компьютерной лингвистики, психологи смещают фокус внимания с языка вообще на само по себе использование слов (word use). Основоположник такого подхода — Джеймс Пеннебейкер — отмечает, что подходы к количественному анализу текстов могут быть разделены на три широкие методологии: метод экспертного тематического контент-анализа (judge-based thematic content analysis), анализ паттернов слов (word pattern analysis) и стратегии подсчета слов (word count strategies) [Pennebaker, Mehl, Niederhoffer, 2002]. В данной работе реализуется стратегия подсчета слов. В современных исследованиях данный подход включает в себя подсчет стандартных грамматических единиц (например, частей речи) или лингвистических параметров, полученных психологически (например,

эмоциональных слова и слов, связанных с достижениями). Этот подход основан на дискуссионном положении о том, что слова, которые мы используем, передают психологическую информацию сверх значения самих этих слов вне зависимости от семантического контекста.

Наиболее часто используемым в стратегии подсчета слов является инструмент, позволяющий анализировать крупные корпуса текстов – the Linguistic Inquiry and Word Count [LIWC; Tausczik, Pennebaker, 2010]. Программное обеспечение было разработано в рамках изучения того, как и почему вербализация негативных переживаний оказывает положительное воздействие на психологическое здоровье [Pennebaker, 1993; Pennebaker, Francis, Booth, 2001]. LIWC выступает как альтернатива структурированным интервью и опросникам и предлагает новый способ проведения исследований. Задачей программы является поиск и подсчет слов, относящихся к психологическим категориям, в наборе текстовых данных. Всего было выделено 80 категорий, каждая из которых прошла валидизацию независимыми экспертами [Pennebaker et al., 2007]. В последней версии программы 2015 года результатом обработки текстового файла являются следующие выходные переменные: количество слов, 4 сводные языковые переменные (аналитическое мышление, влияние, оригинальность текста и эмоциональный тон), 3 категории общих дескрипторов (количество слов в предложении, процент целевых слов, имеющих в словаре, и процент слов в тексте длиннее шести букв), 21 стандартное языковое измерение (например, процент слов в тексте, которые являются местоимениями, артиклями, вспомогательными глаголами и т. д.), 41 словосочетание, затрагивающее психологические конструкции (например, аффект, когниция, биологические процессы, побуждения), 6 категорий личных интересов (например, работа, дом, досуг), 5 неформальных языковых маркеров (например, ругательства) и 12 категорий пунктуации (точки, запятые и т. д.). LIWC – это программа с закрытым словарем. В настоящее время поддерживается большое количество различных языков, в том числе русский.

Исследования с использованием программы LIWC показали, что определенные паттерны использования языка связаны с различными аспектами физического здоровья [Campbell, Pennebaker, 2003] и индикаторами психологического здоровья: например, алекситимии [Meganck et al., 2009] и наличием психиатрического диагноза [Junghaenel, Smyth, Santner, 2008]. Кроме того, были получены свидетельства того, что повествование о травме с меньшим количеством слов когнитивной обработки, большим количеством слов о смерти и большим количеством местоимений первого лица предсказывают увеличение симптомов посттравматического стрессового расстройства через 6 месяцев [Kleim et al., 2018]. Сравнение речи правонарушителей с психопатией и речи других правонарушителей позволило выделить такие лингвистические маркеры психопатов, как большая отстраненность («вы знаете»), более частое использование личных местоимений, меньшее число упоминаний других людей (например, личных имен, семьи) и меньшая эмоциональная выразительность, в том числе низкая частота связанных с тревогой слов [Le et al., 2017].

В этой связи крайне перспективными представляются исследования лингвистической специфичности уровней развития личности. Развитие личности на современном этапе его понимания может быть рассмотрено как процесс постепенного усложнения форм и способов интерпретации личностью себя и действительности [Loevinger, 1976]. Такое понимание создает общность между исследованиями личности и исследованиями языка как с точки зрения теории, так и на эмпирическом уровне.

Возможной операционализацией указанного подхода к пониманию развития личности является уровневая теория развития эго. Методика неоконченных предложений Вашингтонского университета (НПВУ; WUSCT; Washington University Sentence Completion Test), измеряющая восемь диагностируемых уровней развития личности, проявила себя как один из наиболее надежных и валидных проективных инструментов. Вербальный

материал ответов данной методики позволяет проводить анализ средствами компьютерной лингвистики. Руководство к методике НПВУ содержит подробное описание того, как проявляется каждый из уровней развития личности в завершении предложений [Ну, Loevinger, 1996]. Прошедшие обучение инструменту рейтеры (кодировщики) присваивают каждому протоколу один из восьми уровней развития эго. Авторами руководства отмечается, что основа оценки — это характерные для каждого уровня способы выражения респондентом себя через высказывания.

Результаты первой и на сегодняшний день единственной известной попытки применения обсуждаемых методов к изучению проявления уровней развития личности в тексте были опубликованы в 2018 году [Lanning et al., 2018] и сосредоточены на анализе 44 тысяч протоколов НПВУ. Исследование, описанное ниже, опирается на методы, описанные в указанной статье 2018 года. Новизна, предлагаемая в нашем исследовании, заключается в выдвижении дополнительных гипотез, изучении особенностей русскоязычного корпуса и выявлении в нем количественных проявлений уровня развития личности.

Таким образом, *целью* настоящей работы является установление количественных индикаторов уровней развития личности (уровней развития эго) в рамках стратегии подсчета слов.

МЕТОД

Протоколы методики НПВУ были аккумулированы в исследованиях Международной лаборатории позитивной психологии личности и мотивации НИУ ВШЭ, проведенных в период с 2014 по 2019 год. Общая выборка составила 2 081 человек (37 458 завершений предложений). В силу специфики сбора данных, не для всех респондентов оказалось возможным описание демографических показателей. Возраст 1171 респондента составлял от 14 лет до 61 года ($M=21.42$, $SD=7.22$). Параметр пола известен для 1340 респондентов, из которых 309 мужчин и 1031 женщина. Участие всех респондентов было добровольным. Подготовка, анализ и визуализация данных были осуществлены

с использованием компьютерных программ MS Excel 2016, IBM SPSS Statistics 23, Tableau и LIWC2015.

Для определения уровня развития личности была использована 18-пунктная версия НПВУ, апробированная на русскоязычной выборке [Леонтьев, Михайлова, Рассказова, 2010]. В результате процедуры кодирования рэйтер присваивает респонденту один из восьми уровней развития личности (Е2 Импульсивный, Е3 Самозащиты, Е4 Конформизма, Е5 Самосознания, Е6 Совестьности, Е7 Индивидуалистический, Е8 Автономии, Е9 Интеграции). Кодирование протоколов инструмента осуществлялось опытными рэйтерами. При анализе данных учитывалась импрессионистическая оценка по протоколу в целом, а не оценка отдельно по каждому из пунктов методики. Это позволило сохранить разнообразие, свойственное ответам респондентов, т.к. не все ответы одного протокола могут быть отнесены к одному и тому же уровню. Кроме того, при учете данного фактора можно говорить о том, что лингвистические индикаторы наблюдаются у людей на определенном уровне развития личности, а не присущи самому уровню в чистом виде как теоретической модели.

Использование в качестве эмпирической базы исследования корпуса текстов протоколов методики НПВУ позволяет избавиться от такой побочной переменной, как различия в жанрах текста респондентов: отмечается, что жанр влияет не только на тематику используемой лексики, но также на использование функциональных слов и пунктуации [Viber, 1988].

РЕЗУЛЬТАТЫ

Описательные статистики. Оценки протоколов методики НПВУ продемонстрировали распределение выборки по уровням развития личности от Е2 до Е8 (Таблица 1), уровень Е9 Интеграции в используемом в исследовании наборе данных не представлен. Распределение соответствует теоретически ожидаемому для городской выборки: самым распространенным (модальным)

уровнем развития эго оказался уровень Е5 Самосознания.

Таблица 1. Распределение уровней развития личности в изучаемом корпусе

Уровень развития эго	N	Доля (%)
Е2 Импульсивный	4	0,2
Е3 Самозащиты	30	1,4
Е4 Конформизма	397	19,1
Е5 Самосознания	1171	56,3
Е6 Совестью	403	19,4
Е7 Индивидуалистический	70	3,4
Е8 Автономии	6	0,3
Всего	2081	100,0

Дисперсионный анализ синтаксических индикаторов уровня развития личности. В качестве синтаксических индикаторов были выбраны показатели LIWC «Запятыя» и «Кавычки», предположительно отражающие уровень сложности текста и проявляемых автором позиций. На графике средних значений показателя относительно уровней развития личности (Рисунок 1) наблюдается восходящий тренд в использовании запятых, начиная с уровня Е6 Совестью, который сохраняется на уровнях Е7 Индивидуалистическом и Е8 Автономии.

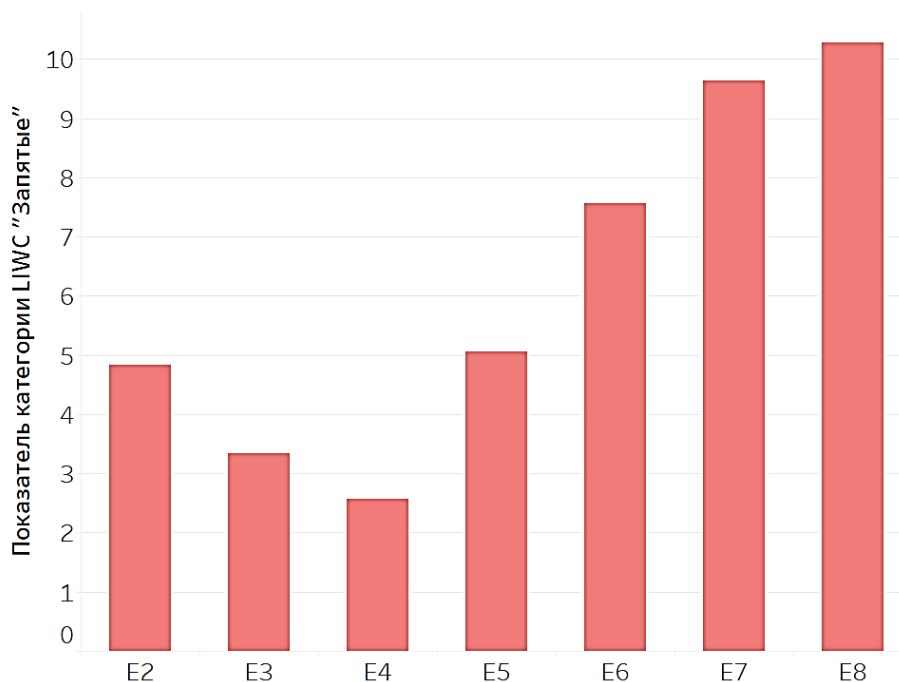


Рисунок 1. Показатели категории LIWC «Запятые» относительно уровней развития личности

Показатели использования кавычек, отраженные на графике (Рисунок 2), характеризуются постоянным возрастающим трендом. Таким образом, визуализация полученных данных позволяет сделать предположение о верности гипотез о возрастании частоты использования запятых и кавычек на высоких уровнях развития личности.

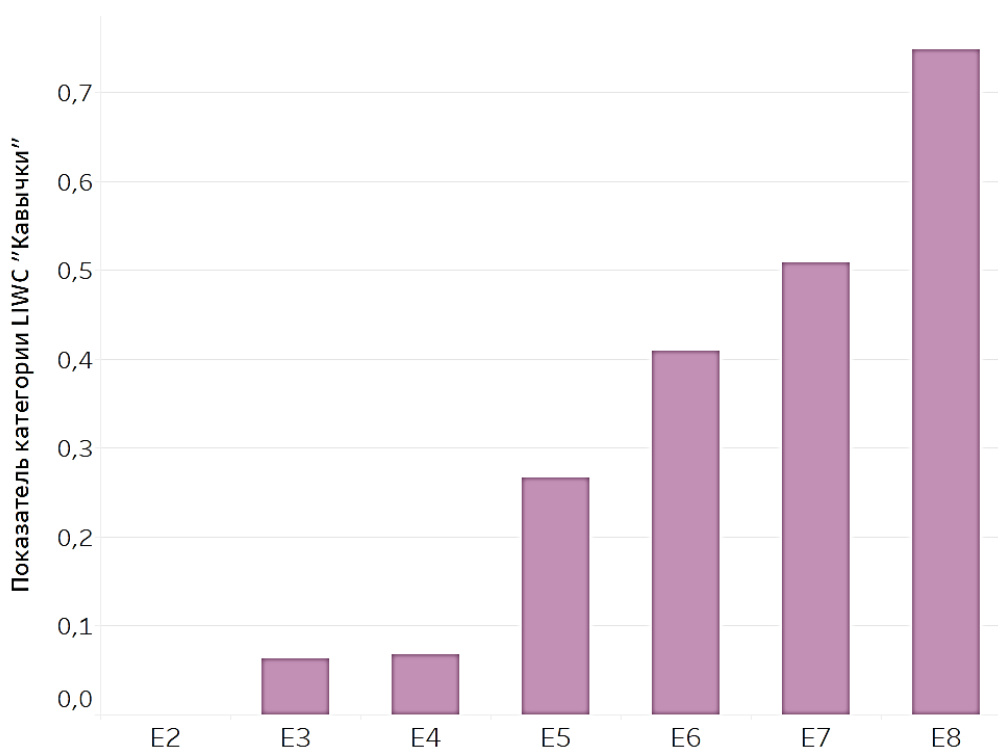


Рисунок 2. Показатели категории LIWC «Кавычки» относительно уровней развития личности

Дальнейшая статистическая проверка гипотезы посредством однофакторного дисперсионного анализа подтвердила значимые различия между уровнями развития эго как для категории «Запятые» [$F(6,2074) = 91.186, p < 0.001$], так и для категории «Кавычки» [$F(6,2074) = 5.939, p < 0.001$].

Дисперсионный анализ лексических индикаторов уровня развития личности. Графики средних значений показывают, что ключевые лексические индикаторы (общее число слов в протоколе, среднее число слов в предложении и процент слов длиннее 6 букв) также увеличиваются по мере повышения уровня развития личности (Рисунки 3 и 4).

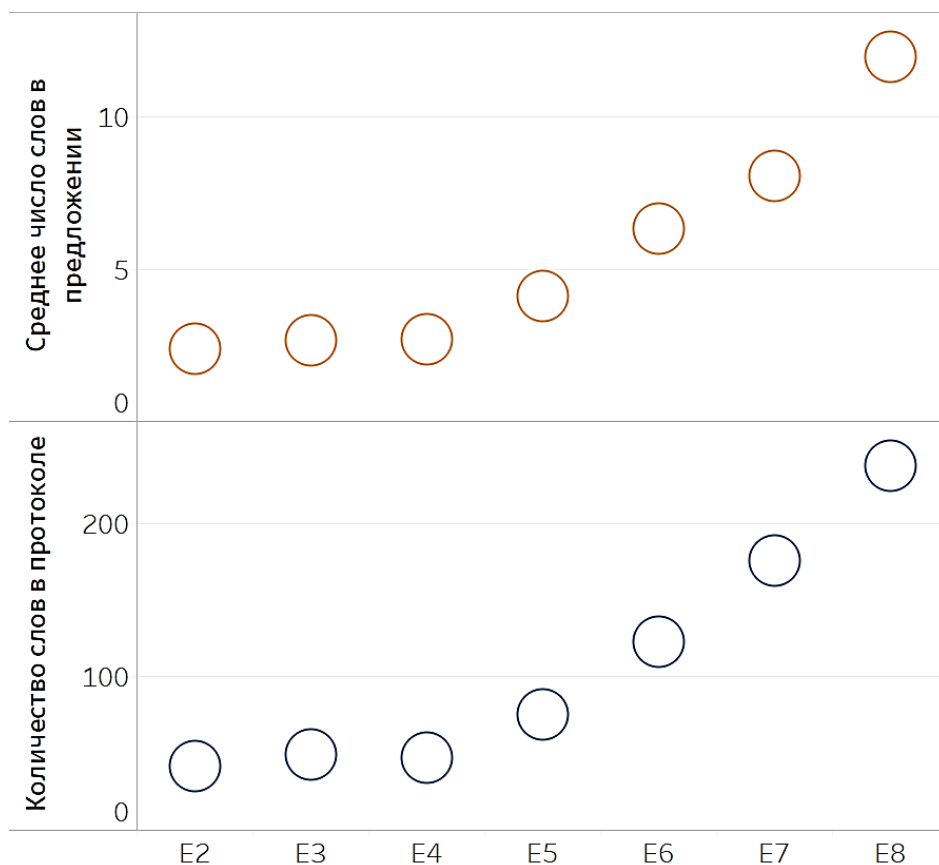


Рисунок 3. Среднее общее количество слов в протоколе и количество слов в предложении относительно уровней развития личности

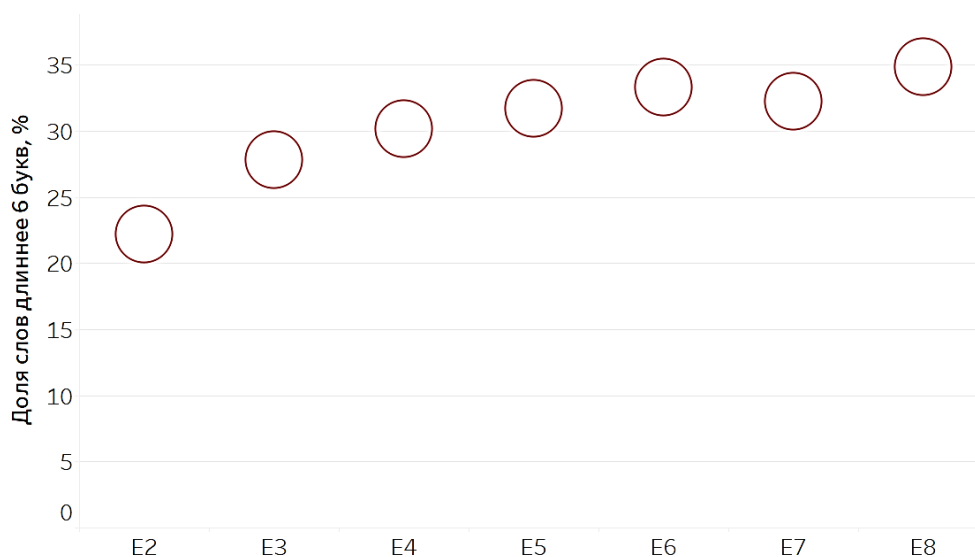


Рисунок 4. Процент слов длиннее 6 букв от общего числа слов относительно уровней развития личности

Как и при анализе синтаксических переменных, для проверки гипотезы о

различиях между уровнями развития эго был применен однофакторный дисперсионный анализ. Результаты показали наличие различий для всех трех категорий — среднее число слов в предложении [$F(6,2074) = 279.223, p < 0.001$], общее количество слов в протоколе [$F(6,2074) = 237.973, p < 0.001$] и доля слов длиннее 6 букв [$F(6,2074) = 10.467, p < 0.001$].

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В предыдущих исследованиях было обнаружено, что более частое использование запятых и кавычек является одними из основных предикторов высокой успеваемости студентов [Robinson, Navea, Ickes, 2012]. Авторы исследования предполагают, что это является проявлением большего интеллектуального взаимодействия с миром. Студенты же с низкой успеваемостью рассматриваются как «живущие в узком и относительно эгоцентрическом мире семьи и друзей, простых идей и простых гедонистических удовольствий» [Ibid], что находит отражение в различных лингвистических параметрах. Схожим образом результаты данного исследования указывают на увеличение использования запятых и кавычек по мере обогащения, усложнения и наполнения разнообразными возможными позициями представлений о себе и о мире в процессе развития личности, что подтверждает поставленные гипотезы.

В исследовании, основанном на анализе англоязычных протоколов НПВУ, было показано, что «по мере развития эго ответы становятся длиннее» [Lanning et al., 2018, p. 328]. Данное исследование подтверждает этот результат на русскоязычной выборке на основании параметров общего количества слов, количества слов в предложении и частоты использования слов, состоящих из более чем 6 букв. Объяснением данной количественной специфики текстов является усложнение понимания себя и связанного с этим нарратива.

ЗАКЛЮЧЕНИЕ

При обращении к синтаксическим индикаторам были обнаружены результаты, расширяющие эмпирические свидетельства, полученные в упомянутой ранее статье. Большое использование запятых на высоких уровнях развития личности свидетельствует об усложнении объяснительных и описательных построений, которые требуют все большего обособления, перечисления и других языковых операций, являющихся функциональным значением использования запятой. Резкое увеличение использования кавычек на высоких уровнях развития личности, вероятно, является косвенным показателем расширения внутреннего многоголосия и диалогичности Я. Полученные результаты являются основанием для подтверждения общей теоретической гипотезы о проявлении уровней развития личности в способах использования слов.

Литература:

1. Леонтьев Д. А., Михайлова Н. А., Рассказова Е. И. Апробация методики незаконченных предложений Вашингтонского университета // Психологическая диагностика. 2010. № 3. С. 6–36.
2. Biber D. Variation across speech and writing. Cambridge University Press, 1988.
3. Campbell R. S., Pennebaker J. W. The Secret Life of Pronouns: Flexibility in writing style and physical health // Psychological Science. 2003. № 1 (14). P. 60–65.
4. Goranson A. et al. Dying Is Unexpectedly Positive // Psychological Science. 2017. № 7 (28). С. 988–999.
5. Hy L. X., Loevinger J. Measuring ego development / L. X. Hy, J. Loevinger, 2nd ed.-e изд., Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc., 1996.

6. Junghaenel D. U., Smyth J. M., Santner L. Linguistic Dimensions of Psychopathology: A Quantitative Analysis // Journal of Social and Clinical Psychology. 2008. № 1 (27). P. 36–55.
7. Kleim B. et al. Early Linguistic Markers of Trauma-Specific Processing Predict Post-trauma Adjustment // Frontiers in psychiatry. 2018. (9). P. 645.
8. Lanning K. et al. Personality development through natural language // Nature Human Behaviour. 2018. № 5 (2). P. 327–334.
9. Le M. T. et al. The Linguistic Output of Psychopathic Offenders During a PCL-R Interview // Criminal Justice and Behavior. 2017. № 4 (44). P. 551–565.
10. Loevinger J. Ego Development: Conceptions and Theories. San-Francisco: Jossey-Bass, 1976.
11. Lyons M., Aksayli N. D., Brewer G. Mental distress and language use: Linguistic analysis of discussion forum posts // Computers in Human Behavior. 2018. (87). P. 207–211.
12. Meganck R. et al. Alexithymia and interpersonal problems: A study of natural language use // Personality and Individual Differences. 2009. № 8 (47). P. 990–995.
13. Pennebaker J. W. et al. The Development and Psychometric Properties of LIWC2007 // LIWC Manual. URL: <http://www.liwc.net/LIWC2007LanguageManual.pdf> (дата обращения: 25.12.2020).
14. Pennebaker J. W. Putting stress into words: Health, linguistic, and therapeutic implications // Behaviour Research and Therapy. 1993. № 6 (31). P. 539–548.
15. Pennebaker J. W., Francis M. E., Booth R. J. Linguistic Inquiry and Word Count: LIWC 2001 / J. W. Pennebaker, M. E. Francis, R. J. Booth, NJ: Lawrence Erlbaum, 2001.
16. Pennebaker J. W., Mehl M. R., Niederhoffer K. G. Psychological Aspects of Natural Language Use: Our Words, Our Selves // Annual Review of Psychology. 2002. № 1 (54). P. 547–577.

17. Qiu L. et al. Personality predicts words in favorite songs // *Journal of Research in Personality*. 2019. (78). P. 25–35.
18. Robinson R. L., Navea R., Ickes W. Predicting Final Course Performance From Students' Written Self-Introductions: A LIWC Analysis // *Journal of Language and Social Psychology*. 2012. № 4 (32). P. 469–479.
19. Tausczik Y. R., Pennebaker J. W. The psychological meaning of words: LIWC and computerized text analysis methods // *Journal of Language and Social Psychology*. 2010. № 1 (29). P. 24–54.
20. Vergani M., Bliuc A. M. The Language of New Terrorism: Differences in Psychological Dimensions of Communication in Dabiq and Inspire // *Journal of Language and Social Psychology*. 2018. № 5 (37). P. 523–540.

References:

1. Biber D. *Variation across speech and writing*. Cambridge University Press, 1988.
2. Campbell R. S., Pennebaker J. W. The Secret Life of Pronouns: Flexibility in writing style and physical health // *Psychological Science*. 2003. № 1 (14). P. 60–65.
3. Goranson A. et al. Dying Is Unexpectedly Positive // *Psychological Science*. 2017. № 7 (28). C. 988–999.
4. Hy L. X., Loevinger J. *Measuring ego development* / L. X. Hy, J. Loevinger, 2nd ed.-e изд., Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc., 1996.
5. Jung haenel D. U., Smyth J. M., Santner L. Linguistic Dimensions of Psychopathology: A Quantitative Analysis // *Journal of Social and Clinical Psychology*. 2008. № 1 (27). P. 36–55.
6. Kleim B. et al. Early Linguistic Markers of Trauma-Specific Processing Predict Post-trauma Adjustment // *Frontiers in psychiatry*. 2018. (9). P. 645.

7. Lanning K. et al. Personality development through natural language // *Nature Human Behaviour*. 2018. № 5 (2). P. 327–334.
8. Le M. T. et al. The Linguistic Output of Psychopathic Offenders During a PCL-R Interview // *Criminal Justice and Behavior*. 2017. № 4 (44). P. 551–565.
9. Leontiev D. A., Mihajlova N. A., Rasskazova E. I. Aprobaciya metodiki nezakonchennyh predlozhenij Vashingtonskogo universiteta // *Psihologicheskaya diagnostika*. 2010. № 3. С. 6–36.
10. Loevinger J. *Ego Development: Conceptions and Theories*. San-Francisco: Jossey-Bass, 1976.
11. Lyons M., Aksayli N. D., Brewer G. Mental distress and language use: Linguistic analysis of discussion forum posts // *Computers in Human Behavior*. 2018. (87). P. 207–211.
12. Meganck R. et al. Alexithymia and interpersonal problems: A study of natural language use // *Personality and Individual Differences*. 2009. № 8 (47). P. 990–995.
13. Pennebaker J. W. et al. The Development and Psychometric Properties of LIWC2007 // *LIWC Manual*. URL: <http://www.liwc.net/LIWC2007LanguageManual.pdf> (дата обращения: 25.12.2020).
14. Pennebaker J. W. Putting stress into words: Health, linguistic, and therapeutic implications // *Behaviour Research and Therapy*. 1993. № 6 (31). P. 539–548.
15. Pennebaker J. W., Francis M. E., Booth R. J. *Linguistic Inquiry and Word Count: LIWC 2001* / J. W. Pennebaker, M. E. Francis, R. J. Booth, NJ: Lawrence Erlbaum, 2001.
16. Pennebaker J. W., Mehl M. R., Niederhoffer K. G. Psychological Aspects of Natural Language Use: Our Words, Our Selves // *Annual Review of Psychology*. 2002. № 1 (54). P. 547–577.
17. Qiu L. et al. Personality predicts words in favorite songs // *Journal of Research in Personality*. 2019. (78). P. 25–35.

18. Robinson R. L., Navea R., Ickes W. Predicting Final Course Performance From Students' Written Self-Introductions: A LIWC Analysis // Journal of Language and Social Psychology. 2012. № 4 (32). P. 469–479.
19. Tausczik Y. R., Pennebaker J. W. The psychological meaning of words: LIWC and computerized text analysis methods // Journal of Language and Social Psychology. 2010. № 1 (29). P. 24–54.
20. Vergani M., Bliuc A. M. The Language of New Terrorism: Differences in Psychological Dimensions of Communication in Dabiq and Inspire // Journal of Language and Social Psychology. 2018. № 5 (37). P. 523–540.