

CLASSIFICATION OF THERAPIST'S INTENTIONS IN PSYCHOTHERAPEUTIC DISCOURSE

*Latynov Vladislav Viktorovich
Vanin Alexander Vladimirovich
Vlasova Arina Sergeevna
Dymova Ekaterina Nikolaevna
Tulyankina Anna Igorevna*

*Federal State Budgetary Institution of Science "Institute of Psychology
of the Russian Academy of Sciences", Moscow*

Abstract. The article describes the System for classifying therapist's intentions in psychotherapeutic discourse developed by the authors. It consists of 10 categories (groups of intentions): "Informing", "Request for information", "General emotional support", "Maintaining positive dynamics of therapy", "Negative evaluation", "Self-disclosure", "Request for building connections", "Interpretation", "Mentoring", "Other". Each category includes smaller units – components/intentions. Using this System, a group of experts analyzed 14 psychotherapeutic sessions. The most frequently identified intentions related to four categories: "Request for information", "Interpretation", "Request for building connections", "General emotional support". The reliability of the System for classifying therapist's intentions (consistency of expert opinions) was assessed using the Fleiss' Kappa coefficient. For most categories (8 out of 10), the level of consistency is assessed as significant. The prospects of using the System for teaching a large language model (ChatGPT) to identify intentions and analyze factors affecting the accuracy of this task are considered.

Keywords: intention, intent analysis, category, component, psychotherapeutic discourse

В данной работе представлена разработанная сотрудниками лаборатории «Технологии искусственного интеллекта в психологии» Института Психологии РАН методика классификации интенций терапевта в психотерапевтическом дискурсе. Разработка этой системы не была самоцелью, а являлась первым шагом на пути решения более сложной задачи – создания модели глубокого обучения на базе больших языковых моделей, позволяющей выявлять и классифицировать интенции в рамках психотерапевтического дискурса.

Основу теоретической модели исследования составила развиваемая в Лаборатории психологии речи и психолингвистики Института Психологии РАН концепция речевых интенций субъекта

и метода их обнаружения – интен-анализа [2; 3; 4]. Неразрывно связанные как с мотивационными и перцептивно-когнитивными процессами субъекта, так и с его включенностью в практическую жизненную деятельность, интенции во многом определяют речевое содержание и его понимание собеседником. Интенции обнаруживаются, с одной стороны, в том, о каких предметах человек высказывается, на что направлено его сознание: на себя, на партнера общения, на окружающую действительность. С другой стороны, важно, как именно об этих предметах говорится [4]. Под речевой интенцией в нашем исследовании понималось намерение терапевта, выраженное им в определенном высказывании.

На первом этапе создания Методики классификации интенций терапевта было проанализировано около тридцати систем аналогичной направленности (как российских, так и зарубежных) [см. напр.: 1; 5; 6; 7]. Поскольку все они имели те или иные ограничения, было принято решение создать собственную оригинальную систему, учтя при этом все лучшее, что было наработано предшественниками.

Методика классификации интенций терапевта состоит из 10 категорий. Категорию можно рассматривать как группу интенций, реализующих одну, достаточно общую направленность терапевта (например, на информирование клиента или интерпретацию его поведения). Каждая категория включает несколько более мелких единиц – компонент/интенций (всего в 10 категориях было выделено 47 компонент). Ниже мы приводим список категорий с краткими их определениями, входящими в них компонентами, а также примерами высказываний терапевта для каждой категории:

1) Информирование (раздел, в котором терапевт сообщает клиенту о различных организационных аспектах терапии, описывает применяемые стратегии, процедуры или программы терапии, рассказывает о закономерностях функционирования психики). Компоненты: организация терапевтического процесса (формальный аспект), организация терапевтического процесса (содержательный аспект), психообразование (описание закономерностей функционирования психики), бытовые факты/универсальное знание.

Пример: «Мы будем работать в рамках когнитивно-поведенческой терапии».

2) Запрос информации (вопросы терапевта, касающиеся событий жизни клиента, его реакций на них, отношений с другими людьми, переживаемых эмоций и желаний). Компоненты: запрос информации о событиях, запрос информации о психической жизни, уточнение и подтверждение полученной от клиента информации. Пример: «Расскажите о ваших отношениях с матерью в тот период».

3) Общая эмоциональная поддержка (раздел, имеющий общий, неспецифический характер, и направленный на создание позитивной и безопасной атмосферы терапии). Компоненты: поддержание контакта, сопереживание, юмор, демонстрация интереса, легитимизация чувств, оправдание критики. Пример: «Я понимаю, как вам было тяжело».

4) Поддержание позитивной динамики терапии (позитивные оценки терапевтом конкретных суждений, действий или качеств клиента, мотивирование и подчеркивание самостоятельности клиента в решении его проблем). Компоненты: одобрение интерпретаций клиента, описание терапевтического эффекта, подчеркивание самостоятельности, позитивная оценка действий или качеств клиента, мотивирование клиента. Пример: «Конечно! Вы абсолютно правы!».

5) Негативное оценивание (раздел, в котором терапевт дает негативную оценку или выражает неодобрение относительно действий, суждений, эмоций, характеристик клиента или других лиц). Компоненты: негативная оценка поведения (внутри сессии), негативная оценка поведения (вне сессии), неодобрение ценностей, негативная оценка интерпретаций клиента, негативная оценка физических и психологических характеристик клиента, угрозы, негативная оценка окружения клиента. Пример: «Я считаю, что Вы не вспомнили то, о чем мы с Вами говорили, а опять поступили слишком импульсивно, несдержанно».

6) Самораскрытие (раздел, в котором терапевт делится своим жизненным опытом, сообщает о возникших у него в ходе терапии

желаниях и эмоциях). Компоненты: желания и чувства терапевта, рассказ о собственных жизненных ситуациях, уход от ответа. Пример: «Мне сложно работать с твоим случаем, много чувств во мне откликается».

7) Запрос на проработку связей (раздел, в котором терапевт просит клиента дать объяснения, интерпретации или прогнозы относительно событий жизни клиента и его психологических реакций). Компоненты: прямой запрос на анализ, запрос на наблюдение, запрос на оценку, запрос на разработку перспективы, запрос прогноза, непрямой запрос на анализ. Пример: «Как Вы думаете, почему Вам сложно делиться своими чувствами даже с самыми близкими?».

8) Интерпретация (раздел, в котором терапевт сообщает о механизмах и закономерностях, касающихся поведения клиента и окружающих его лиц, а также высказывает гипотезы о состоянии, эмоциях и желаниях клиента). Компоненты: экспликация механизмов поведения клиента вне терапии, экспликация механизмов поведения клиента внутри сессии, интерпретация снов, объяснение через метафору и символ, подтверждение точности понимания интерпретации клиента, установление противоречия. Пример: «Именно из-за развода с мужем вы ощущаете себя так неуверенно...».

9) Наставничество (высказывания, в которых терапевт просит клиента совершить определенные действия, либо воздержаться от их выполнения (как в ходе терапевтической сессии, так и вне ее). Компоненты: рекомендации общего плана (с аргументацией), рекомендации общего плана (без аргументации), поведение в конкретной ситуации, придумывание действий, терапевтическая техника, контроль действий. Пример: «Ты должен позаботиться о своей жизни и позволить брату позаботиться о своих делах».

10) Иное (высказывания, не относящиеся ни к одной из представленных выше категорий и, как правило, являющиеся обрывками фраз терапевта, прерванные речью клиента). Компонента: неcodируемое. Пример: «Это совсем...».

На втором этапе разработки Методики классификации интенций терапевта с ее помощью была проведена разметка 14 сессий

(775 реплик терапевта) на русском языке. Эксперты оценивали, какие интенции/компоненты присутствовали в каждой реплике терапевта. При принятии решения они использовали руководство по Методике, включающее определение каждой компоненты и список примеров (от 10 до 15), раскрывающих ее содержание и облегчающих понимание ее смысла. Каждую сессию оценивала команда из трех экспертов, принимающих решения независимо друг от друга.

Всего в 775 репликах терапевта было выделено 1671 компонента/интенция, т. е. примерно две интенции в одной реплике. Наиболее часто выделялись интенции, относящиеся к четырем категориям: «Запрос информации» (462 интенции, 27,6 % от общего количества выделенных интенций), «Интерпретация» (279 интенции, 16,7 %), «Запрос на проработку связей» (262 интенции, 15,6 %), «Общая эмоциональная поддержка» (221 интенция, 13,2 %). Поскольку анализировались так называемые показательные, одноразовые сессии, то выделенный паттерн самых популярных категорий неудивителен: впервые встретившись с клиентом, терапевт в первую очередь стремился создать благоприятную психологическую атмосферу и как можно больше узнать о клиенте.

Кроме того, при помощи коэффициента Fleiss' Карра оценивалась согласованность мнений экспертов по каждой категории. Данная процедура является важным этапом создания любой системы классификации и используется для оценки ее надежности [6]. Значение коэффициента Fleiss' Карра варьируется от -1 (полное несогласие между оценщиками) до 1 (полное согласие между оценщиками). Интерпретация значений Fleiss' Карра: 0,01–0,20: слабое согласие; 0,21–0,40: умеренное согласие; 0,41–0,60: значительное согласие; 0,61–0,80: очень хорошее согласие; 0,81–1,00: почти полное согласие. В целом значения коэффициента Fleiss' Карра для нашей выборки указывают на высокую согласованность оценок экспертов, что свидетельствует о достаточной надежности разработанной Методики классификации интенций терапевта. Для восьми из десяти категорий величина коэффициента Fleiss' Карра находилась в диапазоне от 0,41 до 0,60 (значительное согласие). Лишь

для двух категорий («Негативное оценивание», «Иное») степень согласованности оказалась умеренной (0,21–0,40). Дело в том, что, как правило, негативные оценки в адрес клиента высказывались психотерапевтом в косвенной форме, а это затрудняло их выявление экспертами.

В настоящее время реализуется третий этап исследования, посвященный обучению большой языковой модели (ChatGPT) выявлению интенций терапевта и анализу ее точности в выполнении этой задачи. Планируется провести анализ того, какие категории интенций модель выявляет легче, а какие – труднее. Также будет исследоваться влияние сложности классификации интенций человеком (критерий: рассогласованность оценок экспертов по коэффициенту Fleiss' Карра) на точность их определения моделью.

Литература

1. Кириллова Е.И. Классификация речевых интенций психотерапевта // Психология. Социология. Педагогика. – 2011. – № 11. – С. 52–66.
2. Павлова Н.Д., Афиногенова В.А., Гребенищикова Т.А. Дискурс социальных медиа: интенциональный подход // Психологический журнал. – 2023. – Т. 44. – № 1. – С. 81–90.
3. Павлова Н.Д., Гребенищикова Т.А. Интент-анализ. Основания, процедура, опыт использования. – М.: Изд-во «Институт психологии РАН», 2017. – 151 с.
4. Слово в действии. Интент-анализ политического дискурса / под ред. Т.Н. Ушаковой, Н.Д. Павловой. – СПб.: Алетейя, 2000. – 316 с.
5. Ягнюк К.В. Сессия Карла Роджерса с Глорией: анализ вербальных вмешательств // Журнал практического психолога. – 2014. – № 2. – С. 81–103.
6. Gumz A., Treese B., Marx C., Strauss B., Wendt H. Measuring verbal psychotherapeutic techniques – A systematic review of intervention characteristics and measures // Frontiers in psychology. – 2015. – Vol. 6. – P. 1705.
7. Velasco E.R., Pedersen H.S., Laursen D.H., Skinner T. A meta-narrative review of coding tools for healthcare interactions and their applicability to written communication // PEC innovation. – 2023. – Vol. 3. – P. 100211.