From Words to Measurement: A Transcript-Based Alexithymia Profile (ALI) Aligned with the Perth Alexithymia Questionnaire (PAQ)

Nicholas Venturino

Feelpath

Abstract

Self-report measures of alexithymia, notably the Perth Alexithymia Questionnaire (PAQ, 2018), offer strong psychometrics and broad uptake, yet they inherit structural limits from introspective access, response styles, and literacy. We introduce an interpretable, transcript-based profile—the *Alexithymia Language Index* (ALI)—derived from clients' spontaneous language during psychotherapy. Our initial implementation focuses on extracting structured, verbatim evidence for three PAQ-aligned facets: Difficulty Identifying Feelings (DIF), Difficulty Describing Feelings (DDF), and Externally Oriented Thinking (EOT), including both supporting and contradicting examples for each facet. The current system produces a per-session, per-participant ALI profile consisting of tagged excerpts with brief justifications and deterministic context windows; it does not yet compute numeric facet scores or report validation results. We outline this implementation, the design choices that keep it auditable and clinically readable, and a planned validation program comparing ALI profiles with PAQ-24.

Keywords: alexithymia, psychotherapy process, clinical NLP, PAQ, measurement, psychometrics, interpretability

From Words to Measurement: A Transcript-Based Alexithymia Profile (ALI) Aligned with the Perth Alexithymia Questionnaire (PAQ)

Introduction

Alexithymia—difficulty identifying and describing one's feelings coupled with an externally oriented cognitive style—has substantial implications for psychopathology, interpersonal functioning, and treatment response. It can strain relationships, increase the risk of depression and anxiety, and reduce empathy and engagement in therapy. Fortunately, clinical evidence suggests alexithymia is modifiable: targeted psychological interventions can reduce alexithymia severity and strengthen emotion-identification and emotion-expression skills [Tsubaki and Shimizu, 2024].

But identifying alexithymia—the prerequisite to treatment—remains challenging. The current state-of-the-art diagnostic is the Perth Alexithymia Questionnaire (PAQ), a contemporary, psychometrically robust self-report instrument [Preece et al., 2018]. But self-report instruments rely on introspective access to the patient and are sensitive to response styles and literacy. They also fail to make use of clinically informative patient behavior in psycotherapy sessions.

This paper introduces the *Alexithymia Language Index* (ALI), a transcript-based measure computed from psychotherapy conversations. ALI is designed to (a) align with the PAQ's construct structure (DIF, DDF, EOT; with optional positive/negative splits), (b) satisfy psychometric standards for reliability and validity, (c) remain transparent and auditable, and (d) generalize across settings with explicit fairness testing. We describe the conversational evidence that grounds ALI, the measurement model and scoring, and outline a validation program that positions PAQ-24 as the principal convergent benchmark and will test whether ALI adds predictive value beyond PAQ. Our aim is a practical, time-saving, conversation-derived, behavior-grounded aid that helps clinicians notice and target emotion-language difficulties in session.

Contributions

- A theory-grounded, transcript-based alexithymia profile (ALI) that uses interpretable language markers in place of self-report items while aligning to PAQ's DIF/DDF/EOT structure.

- A minimal, reproducible, and auditable extraction pipeline that turns psychotherapy transcripts into structured, facet-tagged evidence spans (verbatim excerpts with before/after context, facet labels, and brief justifications).

Background and Rationale

Alexithymia structure and the PAQ

The PAQ decomposes alexithymia into *Difficulty Identifying Feelings* (DIF), *Difficulty Describing Feelings* (DDF), and *Externally Oriented Thinking* (EOT), with DIF and DDF further split by affect valence (negative vs. positive). These facets motivate language-based indicators: specificity of emotion labels (DIF/DDF), ability to articulate and link feelings to causes (DDF), and a preference for describing events/objects over internal states (EOT).

PAQ-24 items grouped by facet

The tables below present the 24 PAQ items verbatim, reorganized by facet rather than questionnaire order. This makes the content of each component (DIF, DDF, EOT) more transparent for readers and for mapping to conversation-derived evidence.

Facet (valence)	Item	Wording
Difficulty Identifying Feelings (DIF), negative	2	When I'm feeling bad (feeling an unpleasant emotion), I can't tell whether I'm sad, angry, or scared.
	8	When I'm feeling bad, I can't make sense of those feelings.
	14	When I'm feeling bad, I get confused about what emotion it is.
	20	When I'm feeling bad, I'm puzzled by those feelings.
Difficulty Identifying Feelings (DIF), positive	5	When I'm feeling good (feeling a pleasant emotion), I can't tell whether I'm happy, excited, or amused.
	11	When I'm feeling good, I can't make sense of those feelings.
	17	When I'm feeling good, I get confused about what emotion it is.
	23	When I'm feeling good, I'm puzzled by those feelings.

Table 1

Perth Alexithymia Questionnaire (PAQ-24) items for Difficulty Identifying Feelings (DIF), grouped by valence. Higher ratings indicate higher DIF Preece et al. [2018].

Facet (valence)	Item	Wording
Difficulty Describing Feelings (DDF), negative	1	When I'm feeling bad (feeling an unpleasant emotion), I can't find the right words to describe those feelings.
	7	When I'm feeling bad, I can't talk about those feelings in much depth or detail.
	13	When something bad happens, it's hard for me to put into words how I'm feeling.
	19	When I'm feeling bad, if I try to describe how I'm feeling I don't know what to say.
Difficulty Describing Feelings (DDF), positive	4	When I'm feeling good (feeling a pleasant emotion), I can't find the right words to describe those feelings.
	10	When I'm feeling good, I can't talk about those feelings in much depth or detail.
	16	When something good happens, it's hard for me to put into words how I'm feeling.
	22	When I'm feeling good, if I try to describe how I'm feeling I don't know what to say.

Table 2

Perth Alexithymia Questionnaire (PAQ-24) items for Difficulty Describing Feelings (DDF), grouped by valence. Higher ratings indicate higher DDF Preece et al. [2018].

Facet (valence)	Item	Wording
Externally Oriented Thinking (EOT), general	3	I tend to ignore how I feel.
	6	I prefer to just let my feelings happen in the background, rather than focus on them.
	9	I don't pay attention to my emotions.
	12	Usually, I try to avoid thinking about what I'm feeling.
	15	I prefer to focus on things I can actually see or touch, rather than my emotions.
	18	I don't try to be 'in touch' with my emotions.
	21	It's not important for me to know what I'm feeling.
	24	It's strange for me to think about my emotions.

Table 3

Perth Alexithymia Questionnaire (PAQ-24) items for Externally Oriented Thinking (EOT). Higher ratings indicate higher EOT Preece et al. [2018].

Limits of self-report and opportunity for conversational evidence

Even strong self-reports capture beliefs *about* one's inner life rather than behavior *in* interaction. Psychotherapy transcripts reveal how people actually represent internal states when invited to do so, offering complementary evidence that can be quantified and subjected to the same psychometric scrutiny.

ALI Uses *Different Input* Than Self-Report

ALI is not a questionnaire. Inputs are verbatim therapy transcripts with speaker turns and timestamps. The unit of analysis is (a) *affect-prompt windows*—an affect-focused therapist question and the client's next 1–3 turns—and (b) *spontaneous windows*—client turns that include emotion terms or mental-state verbs without a prior prompt. This preserves the interactional nature of therapy.

Why Not Read PAQ Items Aloud?

Self-report administered in-session reverts to endorsement rather than behavior. ALI captures *spontaneous* and *contextualized* emotion talk and the client's uptake of emotional invitations. This yields a measure closer to real-world functioning and complementary to PAQ rather than a substitute for it.

Method Overview: From Conversation to a Structured ALI Profile

Opportunities to express emotion. Throughout a typical therapy session, the client has opportunities to express emotion. For example, the therapist may ask how the client is feeling after discussing a recent event, or may invite the client to notice what is happening in their body. These moments give concrete chances to see whether feelings are named, described, deflected, or avoided.

Client responses to opportunities. In practice, our current implementation reads the full participant transcript (and the full-session transcript for context) and looks for short, verbatim excerpts that serve as evidence for or against each PAQ facet. We treat direct, specific emotion naming as evidence against DIF/DDF; generic or confused language about feelings as evidence for DIF; difficulty finding words and shallow descriptions as evidence for DDF; and a shift away from inner experience toward tasks, logistics, or external details as evidence for EOT.

Spontaneous affect talk. Outside explicit prompts, we still see meaningful evidence in how clients talk about feelings on their own. The same extraction process can identify excerpts where clients spontaneously name emotions, struggle to describe them, or stay anchored in events and tasks instead of internal states. Over time, this allows us to build up a profile of how often each kind of evidence appears.

Controls (planned). In future work, we intend to add simple controls for transcript length, topic mix, and therapist prompt rate when turning ALI markers into numeric scores. In this paper we focus on the implemented extraction step and do not report numeric scoring.

Illustrative Examples and Client Comparisons

Example Flow with Transcript Excerpt

Therapist: How did that make you feel?

Client: I guess I felt overwhelmed and kind of numb—it's hard to say why.

Therapist: What do you notice in your body when you talk about it?

Client: My chest tightens and I focus on the emails I need to send.

Processing steps (current implementation)

1. **Transcript bundling**. For each session participant, we construct a transcript bundle containing the best available participant transcript (live or cloud) plus the full-session transcript for context.

2. **Facet-wise extraction (LLM)**. For each PAQ facet and valence (DIF negative, DIF positive, DDF negative, DDF positive, and EOT), a language model reads the participant transcript and full-session transcript and returns a structured list of instances. Each instance includes: a short verbatim excerpt drawn from the participant transcript, a brief justification, and a direction flag indicating whether the excerpt supports or contradicts the facet.

3. **Normalization and context**. For every returned instance, we verify that the excerpt exists as a contiguous substring of the participant transcript. When we can find it, we compute fixed before/after context windows deterministically around the excerpt. When we cannot match an excerpt at all, we drop that instance rather than saving unverifiable text.

4. **Profile construction**. All verified instances are flattened into a single ALI profile for that participant and session, with each excerpt labeled by facet (DIF, DDF, EOT), valence (negative, positive, or mixed), direction (supports vs. contradicts), justification, and its before/after context. At present we stop at this profile layer and do not compute numeric scores; planned scoring and validation against PAQ-24 are described separately as future work.

Comparison of Examples and ALI Scores

Example 1. *Therapist:* When you think about last night, what feelings come up?

Client: I don't really know. I had to get to work early and there were a lot of emails. We should probably talk about scheduling next week.

Therapist: Do you notice anything in your body right now?

Client: Not really.

Example 2. *Therapist:* How was that conversation with your sister?

Client: It was fine. I mean, I've got projects piling up and I'm focused on getting things done. Hard to say what I felt about it.

Therapist: Any sensations when you describe it?

Client: Nothing comes to mind.

Example 3. *Therapist:* What's happening for you as you recall it?

Client: I feel sad and a bit anxious because I wanted their approval and didn't get it. My chest feels tight and my stomach drops when I talk about it.

Structured extraction outputs (LLM; compact view)

- **Ex. 1** (prompt window): high provisional ratings on negative DIF items (2, 8, 14, 20; confusion about what the feeling is) and EOT items (3, 6, 9, 15; focus on logistics and tasks), with low endorsement of DDF items.
- **Ex. 2** (prompt window): similar pattern to Ex. 1, with generic affect language ("fine") and sustained attention to projects and scheduling rather than inner experience.
- **Ex. 3** (prompt/spontaneous): lower DIF and EOT ratings (clear naming of feelings and focus on inner experience) and lower DDF (richer description), relative to Ex. 1 and Ex. 2.

The examples above are intended only to show the kind of language we treat as evidence for higher vs. lower DIF, DDF, and EOT. In our current system we store the underlying excerpts, context windows, and facet labels as an ALI profile; future work may derive numeric scores from these markers, but we do not report any such scores here.

Data and Indicators (What We Analyze)

We analyze de-identified psychotherapy transcripts with speaker turns and timestamps.

- **Indicators (computed from text with evidence excerpts):** structured ALI markers, each consisting of a verbatim excerpt from the participant transcript, a brief justification, facet labels (DIF,

DDF, EOT), valence (negative/positive/mixed), direction (supports vs. contradicts the facet), and deterministic before/after context windows.

- **Units:** per-session, per-participant ALI profiles made up of individual markers, rather than global scores.
- **Controls (planned):** in later work, transcript length, lexical diversity/complexity, therapist prompt rate, and topic mix will be considered when turning profiles into numeric summary scores. These controls are not yet implemented in the current system.

Planned Validation (ALI vs. PAQ-24)

- **Question:** Do ALI profiles built from transcript evidence align with the matching PAQ-24 subscales in the way theory predicts (for example, more evidence for DIF in language when PAQ DIF scores are higher)?
- **Data (planned):** Sessions with de-identified transcripts and paired PAQ-24 responses for the same clients.
- **Analysis (planned):** We intend to derive simple numeric descriptors from ALI profiles (for example, counts or weighted summaries of DIF/DDF/EOT markers) and examine how strongly these descriptors are associated with the corresponding PAQ-24 scores across clients (using rank-based correlations such as Spearman) and whether non-matching facets show weaker associations. We do not report any such analyses in this paper.

Planned Reliability Checks

We also plan to assess how stable and trustworthy ALI profiles are once we have sufficient data.

- **Session-to-session stability (planned).** We plan to examine how stable ALI profiles are across adjacent sessions for the same client, using simple test–retest metrics.
- **Human–model agreement (planned).** We plan to create a small, manually coded set of excerpts in which human raters label DIF/DDF/EOT evidence, and to compare those labels with ALI markers using straightforward agreement statistics. These reliability checks are not yet complete.

Challenges We Need to Overcome

We aim for a practical, trustworthy tool and will need to figure out some solutions to the following challenges.

- **Extraction mistakes or hallucinated spans.** Our current system mitigates this by requiring that every excerpt be matched deterministically to the participant transcript and by dropping any instance that cannot be verified. Future work may add a second, tightly scoped repair step for near-misses while maintaining strict verification.
- **Prompt dependence and model drift/versioning.** ALI currently relies on a small set of carefully designed prompts and explicit facet labels. We will need clear versioning and regression tests so that profiles remain comparable over time.
- **Language proficiency and verbosity.** Less fluent or very brief speakers may show fewer opportunities to demonstrate capacity. Any future scoring will need to account for transcript length and language proficiency so as not to over-pathologize sparse speech.
- **Site/data selection bias.** Early ALI profiles will reflect the sites and populations from which transcripts are drawn. We will need to monitor how well the method generalizes beyond the initial sample.
- **LLM choice subjectivity and bias.** Different models and settings may yield different markers. Keeping prompts, models, and post-processing transparent, and regularly auditing outputs, will be important for trust.

Safety and Privacy (Zero-Data-Retention by Design)

- **Processing environment:** On-prem or with a provider under a no-retention data processing addendum; otherwise a local model.
- **De-identification before model calls:** PHI will be masked prior to any LLM use.
- **Zero-retention policy:** Requests use no-training/zero-retention settings.

- **Minimal logs:** Store only hashed window IDs and short evidence spans; never full transcripts.
- **Access and retention:** Role/key-based access, explicit retention windows, deletion on request; incident response is documented.

Conclusion

ALI aims to provide a practical, time-saving, conversation-derived aid that helps clinicians notice and target emotion-language difficulties in session. In this paper we have described an initial implementation that turns psychotherapy transcripts into structured, PAQ-aligned evidence markers for DIF, DDF, and EOT, each grounded in verbatim excerpts with deterministic context. This implementation is a starting point: it is not yet a validated psychometric scale, but rather an auditable layer that can later support scoring and validation against PAQ-24. We see this as a useful foundation for collaborative work with research partners who want transparent, behavior-grounded indicators of alexithymia in conversational data.

Author Contributions

NV conceived the study, led design of the ALI framework, and drafted the manuscript.

Funding

This work was supported by Feelpath. Additional funding sources will be added upon confirmation.

Conflicts of Interest

NV is affiliated with Feelpath, which may integrate ALI into research tooling. Analyses will follow these plans and independent oversight.

Data and Code Availability

Upon acceptance, analysis code and synthetic example transcripts will be released under an open-source license. Access to de-identified research data will follow institutional and regulatory requirements.

References

David A. Preece, Rodrigo Becerra, Ken Robinson, Justine Dandy, and Alfred Allan. The psychometric assessment of alexithymia: Development and validation of the perth alexithymia questionnaire. *Personality and Individual Differences*, 132:32–44, 2018. doi: 10.1016/j.paid.2018.05.011. URL https://doi.org/10.1016/j.paid.2018.05.011.

Kanako Tsubaki and Eiji Shimizu. Psychological treatments for alexithymia: A systematic review. *Behavioral Sciences*, 14(12):1173, 2024. doi: 10.3390/bs14121173. URL https://doi.org/10.3390/bs14121173.

Appendix

Resources and Navigation

- **Code & scoring examples:** `https://github.com/feelpath-research/ali` (versioned prompts, evidence-span extractors).
- **Synthetic transcripts (demo only):** `https://feelpath.com/research/ali/synthetic`.
- **Working scoring manual:** `https://feelpath.com/research/ali/manual`.