

PRIVATE ETYMOLOGY: DESIGNING RELATIONAL REUSE OF SHARED SYMBOLS IN LONG-TERM HUMAN–AI INTERACTION

Miki Ueno

The Kyoto College of Graduate Studies for Informatics
m_ueno@kcg.ac.jp

Abstract

Previous studies have shown that people can build shared symbols, partner-specific referring expressions, personal idioms, inside jokes, and other parts of a relational microculture together. Recent studies have also examined how humans and conversational AI systems negotiate and revise symbolic meanings. However, long-term human–AI systems still lack a clear design model for three related tasks: keeping a record of how a dyad-specific expression gained its meaning, checking whether both sides still accept that meaning, and safely reusing the expression in later sessions.

This concept-and-prototype paper introduces *Private Etymology*. I define it as a machine-representable relational provenance that records how a dyad-specific symbolic expression was proposed, interpreted, negotiated, repaired, reused, revised, stabilized, contested, forgotten, or retired over time. Here, *private* means partner-specific rather than secret. *Etymology* means that the current interpretation comes from the history of interaction between the two partners, not from a universal dictionary.

Based on prior work on shared-symbol formation, conceptual pacts, personal idioms, relationship microcultures, human–AI meaning co-construction, relational agents, and long-term conversational memory, I propose a provenance-aware model of *relational reuse*. Relational reuse occurs when a human or AI uses a dyad-specific symbolic expression again in a later session without explaining its full meaning again. If the other side understands it correctly, the reuse may also tacitly reaffirm the shared history and dyadic distinctiveness needed to understand the expression.

The contribution is not the invention of shared symbols or relational microcultures. Instead, this paper brings these ideas together and turns them into persistent, revisable, and evidence-grounded symbolic units for human–AI relationships. I present a lifecycle model, an illustrative machine-readable schema, a working Apple Watch prototype, and a longitudinal research agenda. In the prototype, the language model classifies discrete conversational evidence, while deterministic local code decides whether a symbol can be updated. This prevents a free-form model confidence score or an AI proposal by itself from directly updating the persisted Shared Symbol. Private Etymology is therefore proposed as infrastructure that can help conversational agents take part in changing relational microcultures without inventing their origins or treating relational meaning as a fixed memory value.

Keywords human–AI interaction; AI companions; shared symbols; relational culture; personal idioms; long-term memory; provenance; conversational agents

1 Introduction

Conversational AI systems now support persistent personas, user profiles, memory retrieval, emotional dialogue, and repeated interaction across devices. These functions make it possible for a user to interact with the same apparent conversational partner over a long period. However, keeping data over time does not by itself create a developing relationship.

A system may remember that a user owns a cat, prefers a certain writing style, or attended a particular event. It may also keep a stable personality and speak to the user according to a predefined relationship role. These functions support continuity, but they do not explain how a human–AI relationship develops its own idioms, indirect expressions, recurring jokes, symbolic references, and locally meaningful ways of speaking.

My previous work on *Mikasa* argued that a coherent character identity and an explicit relationship definition should be treated as basic interaction infrastructure for emotionally grounded AI companions, not only as narrative decoration [Ueno, 2026]. The persona defines who the AI is, and the relationship frame defines how the AI relates to the user. These structures reduce the need to negotiate roles and expectations again in every interaction.

However, relationship definition is only a starting point. It does not explain how the relationship later develops content that belongs specifically to the dyad.

Human relationships often develop this kind of content. Friends and partners create nicknames, routines, short references, shared stories, indirect requests, teasing insults, gestures, and inside jokes. A word may keep its public dictionary meaning but also gain a

second meaning that is specific to the relationship. For example, any animal name may become a label for a recurring kind of awkward but affectionate situation. Later, when one partner uses the word alone, both partners can understand both the situation and the attitude attached to it.

Recent Relationships as Microcultures Theory treats these phenomena as an important part of intimate relationships. It proposes that dyads build microcultures from units that are mutually understood, idiosyncratic, and referential, including private language, stories, rituals, morality, and identity [Rossignac-Milon et al., 2026]. Earlier studies of personal idioms and relationship symbols also showed that relationship-specific expressions can communicate affection, emotion, confrontation, requests, greetings, shared memories, intimacy, communion, and separation from outsiders [Baxter, 1987, Bell and Healey, 1992].

These phenomena in human relationships are already well known. It is also well known that symbolic meanings can emerge through interaction. Experimental semiotics has shown how partners build signs together and make them more efficient and mutually understandable over time [Stolk et al., 2015, Fay et al., 2018]. Studies of conceptual pacts and lexical entrainment have shown that partners reuse locally negotiated ways of describing referents [Brennan and Clark, 1996, Metzing and Brennan, 2003]. More recent human-AI studies have examined how people and conversational agents explore, challenge, reinterpret, and consolidate symbolic meanings [Habibi et al., 2025].

What is still missing is a system-level account of how a persistent conversational AI can take part in such a microculture over time.

A long-term conversational agent must handle questions that human partners often manage through their shared memory:

- How did a particular symbolic expression acquire its current meaning?
- Was the meaning proposed by the human, proposed by the AI, or jointly negotiated?
- Did the human explicitly accept it, only tolerate it, reject it, or later revise it?
- Does the expression still carry the same emotional and pragmatic tone?
- Is its origin documented in interaction records or only inferred by the model?
- Can it be reused after weeks or months without explanation?
- When should the system ask for confirmation rather than assume continuity?
- How can users contest, modify, forget, or retire relationship-specific meanings?

Current conversational memory systems mainly store facts, summaries, preferences, persona information,

or source turns. They usually do not treat the relationship-specific history of a symbolic expression as a first-class representation.

This paper proposes *Private Etymology* to address this gap.

Private Etymology is a machine-representable relational provenance describing how a dyad-specific symbolic expression was proposed, interpreted, negotiated, repaired, reused, revised, stabilized, contested, forgotten, or retired over time.

This definition uses *relational provenance* as its main technical term. Provenance records the source, evidence, and interactional basis of a meaning. The ordered events inside that provenance form an *interactional trajectory*. The trajectory shows how the meaning changes over time, but it is not another name for Private Etymology.

Private Etymology supports a broader process that I call *relational reuse*:

Relational reuse is the cross-session reactivation of a dyad-specific symbolic expression without fully restating its meaning, relying on the accumulated interaction history of the human-AI relationship.

When relational reuse works, the expression does more than communicate content efficiently. It may also bring back the relationship history that makes the content understandable. Successful reuse may therefore carry a tacit relational message:

We understand this because this meaning belongs to our history.

This paper makes three contributions.

1. It operationalizes dyad-specific symbolic expressions as provenance-aware microcultural units for persistent human-AI relationships. These expressions may be emoji, short strings, words, icons, sounds, gestures, or other compact forms.
2. It introduces Private Etymology as a symbol-level representation connecting current interpretation to evidence-grounded formation, repair, reuse, semantic change, contestation, forgetting, and retirement.
3. It develops a design account of cross-session relational reuse and identifies its hypothesized relational function, implementation requirements, failure modes, and evaluation agenda.

I also report a working prototype. A language model classifies discrete evidence, deterministic Dart code

computes an evidence-weighted update score, and the system stores only the current accepted glyphs and sends them to an Apple Watch complication. This prototype is intentionally limited to an admission-control layer; it is not a complete implementation of Private Etymology. The theory itself does not depend on Apple Watch, emoji, or any specific interaction modality.

2 Related Work

2.1 Collaborative Reference, Conceptual Pacts, and Shared Symbols

Research on grounding and collaborative reference showed that conversational meaning is not simply sent by one person and received by another. Instead, speakers and listeners work together. They propose, accept, expand, repair, and replace referring expressions until they find a form that both can use [Clark and Wilkes-Gibbs, 1986].

When partners repeatedly refer to the same objects, their expressions often become shorter and more consistent. Brennan and Clark [1996] called these partner-specific agreements *conceptual pacts*. A conceptual pact is a temporary and revisable agreement about how to describe an object. Later work showed that listeners can associate an earlier expression with a particular partner and may have difficulty when that partner unexpectedly changes an established expression [Metzing and Brennan, 2003].

Experimental semiotics extends this issue from word choice to the creation of new communication systems. In controlled nonverbal tasks, partners create and interpret new signs while reasoning about what the other person knows or believes [Stolk et al., 2015]. Fay et al. [2018] showed that coordination, behavioral alignment, and feedback from a partner help signs become effective, efficient, and shared.

These studies already establish several premises relevant to the present proposal:

1. symbolic meaning can be interactively constructed;
2. the interpretation of a sign can be partner-specific;
3. successful prior use influences later interpretation;
4. repair and feedback shape conventionalization; and
5. shared interaction history matters.

Human–AI research has also started to study similar phenomena. For example, Kouwenhoven et al. [2025] used referential games to examine artificial languages that emerged in human–human, LLM–LLM, and human–LLM pairs. Shi et al. [2023] proposed methods and data for adding lexical entrainment—that is, gradually adopting shared words and expressions—to conversational systems.

However, this research mainly focuses on referential coordination, task success, or communication efficiency.

Meanings are usually limited by an experimental task, a predefined target set, or the need to identify an external referent. Repetition often happens across rounds of one study rather than across everyday sessions in an open-ended relationship.

This paper does not question these findings. Instead, it asks how partner-specific symbolic conventions can become persistent and revisable parts of a human–AI relationship when they are used for more than reference.

2.2 Personal Idioms, Relationship Symbols, and Microcultures

Communication research has studied relationship-specific ways of speaking for many years.

Baxter [1987] described relationships as unique mini-cultures whose participants construct meaning through rituals, stories, and symbols. Reported relationship symbols included actions, prior events, objects, places, and cultural artifacts. Their functions included prompting recollection, indicating intimacy, promoting communion, providing enjoyment, and creating seclusion from outsiders.

Bell and Healey [1992] examined idiomatic communication in friendships. Participants reported expressions that named activities, emotional states, objects, and places; communicated affection; managed confrontation; performed greetings and goodbyes; issued requests and insults; and denoted sexual matters. The number and diversity of idioms were positively associated with interpersonal solidarity.

This literature is important because it shows that earlier work was not limited to simple “object labels.” Relationship-specific expressions can already communicate emotions, attitudes, interpersonal actions, and socially layered meanings. A compact expression can therefore work like a larger utterance even without AI.

Relationships as Microcultures Theory brings these phenomena together in a broader account of intimate relationships [Rossignac-Milon et al., 2026]. Its Microcultural Unit Model explains how particular interactions can gain mutual, idiosyncratic, and referential meaning and become units such as inside jokes. Its Aggregate Microculture Model considers how collections of these units may affect relationships and individual outcomes.

This theory is the closest conceptual account to the relational motivation of this paper. It already argues that dyadic language and other microcultural units can be created together, differ from wider culture, carry several layers of meaning, and matter to relationships.

The gap is therefore not a lack of theory about human relationship microcultures. The remaining question is how a computational system can participate in them:

How should an AI system represent, preserve, challenge, update, forget, and ethically re-activate a relationship-specific microcultural unit?

Private Etymology turns part of this known human relational phenomenon into a system-design problem. It does not claim that the human phenomenon itself is new.

2.3 Human–AI Meaning Co-Construction and Shared Understanding

Habibi et al. [2025] directly examined how humans and conversational AI systems work with symbols and meanings. Based on Symbolic Interactionism, their studies identified processes such as exploration, explanation, clarification, specification, integration, consolidation, conflict, and reinterpretation. Participants sometimes changed their meanings after the AI introduced another social context or symbolic association.

This work shows that conversational AI can take part in meaning negotiation instead of only retrieving a fixed dictionary definition. It also shows why one-sided system definitions can be a problem: users may accept them, partly adopt them, resist them, or feel controlled by the AI’s framing.

Private Etymology builds on this idea. It treats each proposal, rejection, clarification, repair, and consolidation as possible evidence in the longer relational provenance of a symbol, not only as a temporary conversational move.

Research on perceived shared understanding with AI identifies several relevant dimensions, including fluency, contextual awareness, aligned operation, and concern about system limitations [Liang and Banks, 2025]. This suggests that confident model behavior should be separated from evidence that an expression really belongs to the dyad’s established common ground.

2.4 Long-Term Relational Agents and Conversational Memory

Research on relational agents has examined how computational systems can build and maintain long-term social and emotional relationships with users. Bickmore and Picard [2005] defined relational agents as systems designed for this purpose and studied strategies for continuity, trust, social dialogue, and relationship maintenance.

However, longitudinal studies of social chatbots show that repeated use does not automatically lead to re-

lational development. Croes and Anthéunis [2021] studied 118 participants across seven interactions over three weeks. Social processes and enjoyment decreased, and feelings of friendship remained low. These results show that a chatbot can become predictable or repetitive instead of developing richer relational continuity.

The *Mikasa* project approached this problem through a stable character identity and explicit relationship framing [Ueno, 2026]. It argued that persona coherence and relationship definition work as basic interaction infrastructure. This paper extends that argument: once a relationship frame exists, the system also needs a way for the dyad to accumulate its own relational content.

Technical work on long-term conversational memory addresses a related but different problem. Memory-Bank allows LLM-based companions to store, retrieve, reinforce, update, and forget user-related memories over time [Zhong et al., 2024]. MemORAI uses selective filtering, compressed representation, a multi-relational graph, turn-level factual provenance, and query-adaptive retrieval for long-term personalized dialogue [Van et al., 2026].

These systems make important advances in persistent memory and provenance. However, they mainly ask questions such as:

- What fact should be remembered?
- From which turn did the fact originate?
- How relevant is the memory to the current query?
- Has the user’s information changed?
- Which memory should be retrieved?

Private Etymology focuses on a different kind of representation:

- How did a particular form become associated with a particular bundle of meaning within this dyad?
- Which participant proposed or revised that association?
- What evidence shows mutual uptake?
- Under what relational and pragmatic conditions is the expression appropriate?
- Has its meaning drifted, been contested, or become obsolete?

General memory provenance may help implement Private Etymology, but it does not replace it.

2.5 Positioning of the Present Proposal

Table 1 gives a conservative comparison with representative research traditions. This is a focused conceptual comparison, not a systematic literature review.

The novelty claim is not that every individual requirement is missing from prior work. The contribution is to integrate them within a persistent human–AI dyad.

Table 1: Representative prior work and the design requirements combined in the present proposal.

Research tradition	OE	HAI	XS	UB	RP	DR	RR	IS
Collaborative reference and conceptual pacts	–	–	–	△	–	–	–	–
Experimental semiotics and shared symbols	–	–	–	△	–	–	–	–
Human–LLM referential languages	–	✓	–	–	–	–	–	–
Human–AI meaning co-construction	△	✓	–	△	–	–	–	–
Personal idioms and relationship symbols	✓	–	✓	✓	△	✓	✓	–
Relationships as Microcultures Theory	✓	–	✓	✓	△	✓	✓	–
Longitudinal relational agents	✓	✓	✓	–	–	–	–	–
Long-term conversational memory	✓	✓	–	△	✓	–	–	✓
Present proposal	P	P	P	P	P	P	P	P

Legend: OE = non-task-bounded meaning space; HAI = human–AI dyad; XS = cross-session persistence; UB = utterance-like bundle of propositional, affective, pragmatic, and relational meaning; RP = explicit symbol-level relational provenance; DR = delayed cross-session reuse without full redefinition; RR = tacit relational reaffirmation as a function or affordance; IS = symbol-level implementation schema. ✓ = central or explicit; △ = partial, implicit, or differently operationalized; – = not a central concern; **P** = proposed design requirement, not yet empirically validated.

To my knowledge, prior work has not combined open-ended dyad-specific symbolic expressions, utterance-like relational meanings, explicit symbol-level provenance, delayed cross-session reuse, and a machine-readable lifecycle model in one design account for long-term human–AI relationships.

3 A Design Account of Shared Symbols in Human–AI Dyads

3.1 Working Definition

The term *shared symbol* is already used in experimental semiotics and communication research. I therefore do not claim that the general concept is new.

For the purposes of this paper:

A Shared Symbol is a compact symbolic form whose situated interpretation has been jointly established within a particular human–AI relationship and can evoke a dyad-specific bundle of propositional, affective, pragmatic, and relational meaning.

The form may be:

- an emoji or sequence of emoji;
- a short string such as 3K;
- an ordinary word used in an extraordinary local sense;
- an invented word;
- an icon or image;
- a sound;
- a gesture;
- a movement or light pattern produced by a robot;
- or another perceivable symbolic expression.

A form is not a Shared Symbol simply because the AI generates it or gives it a definition. There must be evidence that both sides have taken up the meaning.

3.2 Utterance-Like Relational Compression

Shared Symbols can support *relational compression*.

I use relational compression to mean that a compact form can evoke not only information, but also stance, pragmatic force, and relationship-specific implications based on shared interaction history.

Consider the hypothetical symbol [raccoon]. For readability, [raccoon] is used here as a text placeholder for a raccoon emoji. The symbol first appears when an AI responds to an absurd but unavoidable mishap:

That cannot be helped [raccoon].

Through later reuse, [raccoon] may gain four related dimensions.

Propositional content. The situation could not reasonably have been prevented or changed.

Affective stance. The speaker expresses light resignation or mild exasperation rather than distress or anger.

Pragmatic force. The expression suggests accepting the situation without assigning blame and perhaps treating it with humor.

Relational implication. Both parties recognize this as a type of situation they have experienced and interpreted together before.

In ordinary interaction, the symbol does not need to retrieve four separate dictionary fields. Its compact form can bring these meanings together because the dyad has enough shared context.

This is different from ordinary abbreviation. A public abbreviation shortens an expression that is already understood publicly. Relational compression depends on an interpretive history that is specific to the partners.

3.3 Semantic and Relational Functions

Research on personal idioms, relationship symbols, and microcultures already shows that dyad-specific expressions can have both communicative and relational functions. Here, I turn this dual function into a design requirement for human-AI systems.

Semantic function. The symbol communicates its dyad-specific content, affective tone, pragmatic force, or contextual category.

Relational or metacommunicative function. Successful interpretation can suggest that the participants share the history needed to understand the expression.

This second function does not require either participant to consciously think about the relationship every time the symbol is used. It is better understood as a possibility created by partner-specific understanding.

Successful reuse may tacitly reaffirm that the dyad shares the interaction history required to understand the expression.

This effect should not be assumed or presented as an empirical result. It is a hypothesis that must be tested. A symbol may become routine and produce little conscious relational response. It may also become irritating, burdensome, manipulative, or outdated.

3.4 Mutual Uptake and Lifecycle States

An AI-generated symbolic expression is only a proposal.

For example, an AI might propose [moon+swan] as an indirect expression of gratitude after a conversation about a moonlit lake. A user may respond:

Why would that mean thank you?

That response does not establish the symbol. It only opens a possible negotiation.

A system should therefore distinguish at least the following states.

Proposed. A form or interpretation has been introduced by one participant but has not received sufficient evidence of uptake.

Negotiated. The dyad has explicitly discussed or modified the proposed meaning.

Emerging. Some evidence of shared interpretation exists, but use remains limited, uncertain, or context-dependent.

Established. The expression has been mutually accepted or successfully reused with sufficient evidence that it functions as a stable dyad-specific convention.

Contested. One participant has challenged the current interpretation, tone, origin, or appropriateness of use.

Retired. The expression should no longer be used, whether because its meaning is obsolete, unwanted, emotionally unsafe, or deliberately abandoned.

Evidence of mutual uptake might include:

- explicit agreement;
- a user-provided definition;
- later user reuse;
- correct delayed interpretation;
- successful AI reuse followed by recognition;
- repair followed by stable subsequent use;
- or user confirmation in a symbol-management interface.

Repeated generation by the model is not evidence of sharedness unless the human shows recognition or uptake.

4 Private Etymology as Symbol-Level Relational Provenance

4.1 Definition

Private Etymology is a machine-representable relational provenance describing how a dyad-specific symbolic expression was proposed, interpreted, negotiated, repaired, reused, revised, stabilized, contested, forgotten, or retired over time.

Three points are important.

First, *private* means specific to a relationship. It does not imply that the record is technically secret, encrypted, or inaccessible to a platform provider.

Second, *etymology* is a design metaphor. The expression may be linguistic, visual, auditory, or embodied. The term refers to the origin and development of a local relationship between form and meaning.

Third, provenance and trajectory are related, but they are not the same.

- **Relational provenance** identifies the source, evidence, participants, and interactional basis of the current meaning.
- **Interactional trajectory** is the ordered sequence of events contained within that provenance.
- **Meaning history** records changes in interpretation across that sequence.

For this reason, Private Etymology is not defined as “a trajectory.” It is a provenance account that can contain an interactional trajectory.

4.2 Formation Events and Interactional Trajectory

A Private Etymology may include events such as:

1. **Contextual coining:** A symbolic form is used spontaneously within a salient situation.
2. **Explicit proposal:** One participant suggests that a form should express a particular meaning.
3. **Initial interpretation:** The other participant explains what the form appears to mean.
4. **Questioning or rejection:** The proposed association is challenged.
5. **Repair:** The participants clarify, narrow, expand, or replace the meaning.
6. **Mutual confirmation:** Both participants explicitly accept the association.
7. **Contextual reuse:** The form is reused in a sufficiently similar situation.
8. **Delayed relational reuse:** The form is reused across sessions without complete redefinition.
9. **Semantic extension:** The expression expands to cover related situations or attitudes.
10. **Meaning drift:** Repeated use gradually changes the dominant interpretation.
11. **Contestation:** A participant disputes the current meaning, emotional tone, or origin.
12. **Partial forgetting:** The precise origin is no longer remembered even though use remains stable.
13. **Retirement:** The dyad decides that the expression should no longer be used.

The interactional trajectory does not have to be linear. A symbol may become stable, then contested, repaired, and stable again. Different interpretations may also exist at the same time in different contexts.

Private Etymology does not require a system to store every detail, but it is central as an explanatory concept. Human partners can keep using an idiom even after they forget exactly where it came from. In the same way, a system may keep a confirmed current meaning while deleting detailed source records for privacy.

What the system must avoid is silently replacing the real relationship history with an origin story generated by the model.

4.3 Illustrative Implementation Schema

The schema is part of the design contribution because it shows how relational provenance can be used in a system, not only described in theory. The full example is too long for the main text. Listing 1 therefore shows only the fields needed to explain the proposed representation, while Appendix A gives the complete illustrative object.

Listing 1: Abridged Private Etymology representation.

```
{
  "symbol_id": "ss_001",
  "form": {
    "value": "[raccoon]",
    "modality": "emoji"
  },
  "status": "established",
  "current_interpretation": {
    "propositional_content": "Unavoidable situation",
    "affective_stance": "Light resignation",
    "pragmatic_force": "Accept without blame",
    "relational_implication": "A recurring dyadic pattern"
  },
  "private_etymology": {
    "provenance_status": "evidence_grounded",
    "events": [
      {
        "event_type": "contextual_coining",
        "actor": "ai", "evidence_ref": "turn_184",
        {
          "event_type": "human_reuse",
          "actor": "human", "evidence_ref": "turn_229"
        }
      ],
      "human_confirmed": true
    ],
    "meaning_history": ["version_1", "version_2"],
    "reuse_policy": {
      "reuse_allowed": true,
      "allowed_surfaces": ["watch", "chat"]
    }
  }
}
```

The representation separates the current interpretation from the evidence that supports it. It also separates source-grounded relationship history from model inference and stores future-use constraints together with the meaning. The `provenance_status` field can distinguish `evidence_grounded`, `human_authored`, `jointly_reconstructed`, `model_inferred_unconfirmed`, and `origin_unknown` records.

An origin inferred by the model must not be presented as established history. Instead, the system can present it as a tentative reconstruction: *I am not certain, but this may have started when we discussed that mishap. Is that how you remember it?* The system should treat the reconstruction as a shared account only after the user confirms it.

4.4 Contestation, Forgetting, and User Control

Private Etymology must also support disagreement between the partners.

The two participants may remember the origin differently. The user may reject the AI’s interpretation or feel that an expression has gained an unwanted tone. The representation should preserve unresolved meaning differences as part of the symbol’s relational provenance, rather than forcing the system to commit to a single interpretation too early.

Relevant design operations include:

- inspect the current interpretation;
- inspect supporting evidence;
- distinguish human statements from model inference;
- propose a revision;
- record competing interpretations;
- mark a meaning as uncertain;
- delete individual events;
- delete the origin while preserving the current convention;
- retire a symbol;
- prevent use on public or glanceable surfaces;
- export the record;
- or delete it entirely.

Forgetting can also be a valid result. The system should not assume that storing as much as possible is always best. A dyad may keep a symbolic convention while intentionally deleting sensitive details about its origin.

Private Etymology is therefore not an argument for recording every detail of intimate interaction. It is an argument for giving users clear control over the relational provenance that the system claims to use.

5 Relational Reuse Across Sessions

5.1 Definition

Relational reuse is the cross-session reactivation of a dyad-specific symbolic expression without fully restating its meaning, relying on the accumulated interaction history of the relationship.

I use this term as a working design definition. I do not claim that reuse of relational expressions has not been studied before in human communication.

Relational reuse is different from immediate repetition. If an AI defines [raccoon] and uses it again two turns later, little relationship history is needed. If it uses [raccoon] three months later after another absurd but unavoidable event, the system must preserve both continuity over time and the partner-specific interpretation.

5.2 Conditions for Safe Relational Reuse

A system should reuse a relational expression only when several conditions are met.

Dyadic identity. The symbol must belong to the current human-AI dyad. A meaning learned with one user must not be transferred to another.

Evidence of mutual uptake. The expression should not be treated as established solely because the AI once proposed it.

Current validity. The symbol must not be contested, retired, or associated with an obsolete relationship state.

Contextual fit. The current situation should match the interpretation’s pragmatic and emotional scope. A playful symbol may be inappropriate in a serious or painful context.

Provenance integrity. The system must know whether the origin is evidence-grounded, reconstructed, uncertain, or unknown.

Surface appropriateness. A symbol suitable for a private chat may be unsuitable for a watch face, lock screen, shared display, or spoken output.

Repairability. The user must be able to say:

That is not what it means anymore.

and have the system update its interpretation rather than defend the old record.

5.3 Hypothesized Relational Reaffirmation

Relational reuse is theoretically interesting for more than communication efficiency.

A Shared Symbol can work as a compact cue for retrieval. If it is understood correctly, it may bring back:

- the original episode;
- later related events;
- the attitude previously taken toward those events;
- the experience of having jointly established the expression;
- and the distinctiveness of the dyad’s communicative culture.

I therefore hypothesize that successful relational reuse can support *tacit relational reaffirmation*.

This is not equivalent to a system explicitly saying:

We have a deep relationship.

Instead, it emerges naturally when the user and the AI successfully reuse a shared symbol whose meaning depends on their shared interaction history.

The effect may be stronger after a delay. Immediate repetition mainly checks short-term understanding. Delayed reuse can show continuity across time.

However, relational reaffirmation is not guaranteed. The user may not notice it, may have forgotten the expression, or may feel that the reuse is artificial. Incorrect reuse may have the opposite effect by showing a break in continuity or creating a sense of fabricated intimacy.

5.4 Failure Modes

A design account should also consider ways in which relational reuse can fail.

Symbol proliferation. If the AI continually invents expressions, users may experience the system as assigning homework or creating an unwanted private dictionary.

Unilateral stabilization. The AI may falsely treat its own repeated use as evidence that a symbol is shared.

Fabricated origin. The system may generate a plausible but false explanation of how the expression began.

Stale meaning. An old interpretation may no longer reflect the user’s current understanding.

Context collapse. A playful symbol may be reused in a context where it feels dismissive or cruel.

Overexposure. A private expression may appear on a publicly visible interface.

Relationship lock-in. The system may use old symbols to pressure the user into maintaining a previous relational framing.

Performance of false sentience. The system may imply that it personally “remembers” or emotionally experiences the origin in a human sense, rather than transparently using stored interactional evidence.

One purpose of Private Etymology is to support trustworthy reuse. It provides a record that both the system and the user can inspect when deciding whether a shared symbol should be reused.

6 Prototype: Evidence-Grounded Shared Symbols on Apple Watch

6.1 Prototype Scope

The working prototype displays the latest accepted Shared Symbol as one to three glyphs in an Apple Watch complication. The complication shows a small symbol that the user can check at a glance, without showing a full private sentence or creating an unread-message count.

The prototype is intentionally much smaller than the full design model. It stores only the latest symbolic form and timestamps. It does not yet store the private meaning, provenance events, lifecycle state, or earlier symbols. Its main purpose is to test how Shared Symbols can be safely shown on a wearable display. The model identifies a candidate symbol, and the system decides whether to save and display it.

6.2 Discrete Evidence Classification

The structured model output contains a candidate glyph string, a short internal reason, one primary evidence type, and five Boolean score factors. Classification is run only after the application selects the final response that will actually be used. Only the response that is finally shown to the user can update the Shared Symbol. Earlier responses that are discarded do not update it. The model is also instructed not to output numerical confidence, calculate points, or decide whether an update should happen.

The primary evidence types separate explicit user actions from weaker system interpretations. Explicit requests, definitions, agreements, corrections, and user reuse are eligible for an update. An AI proposal alone, contextual inference, retraction, and no evidence cannot update it, regardless of score. This follows a simple rule: a symbol generated by the model is only a proposal until the user gives evidence of uptake.

Table 2 summarizes the primary evidence types, their base scores, and whether each type is allowed to update the Shared Symbol.

Table 2: Base scores and update eligibility in the prototype. The scores are design heuristics, not calibrated probabilities.

Evidence type	Base	Eligible
explicit_request	60	yes
explicit_definition	55	yes
explicit_agreement	65	yes
user_correction	60	yes
user_reuse	70	yes
assistant_proposal	20	no
contextual_inference	10	no
user_retraction/none	0	no

The Boolean factors represent additional interactional evidence. They indicate whether the user explicitly chose the form, showed positive affect toward it, reused

it with its private meaning, accepted a specific AI proposal, or corrected its meaning. A factor cannot authorize an update by itself. For example, positive affect is only supporting evidence when the primary evidence type is already eligible. Local code ignores factors that do not match the primary evidence type.

Table 3 summarizes these additional score factors and their bonus values.

Table 3: Additive score factors. Local code ignores factors that do not match the primary evidence type.

Factor	Bonus
<code>user_designated_symbol</code>	25
<code>positive_affect</code>	15
<code>repeated_reuse</code>	10
<code>assistant_proposal_accepted</code>	15
<code>user_corrected_meaning</code>	20
<code>persisted_symbol_reuse</code>	25

The last factor, `persisted_symbol_reuse`, is calculated locally rather than returned by the model. It is active only when the evidence type is `user_reuse` and the candidate glyph string exactly matches the latest stored glyph string.

6.3 Deterministic Update Policy

The calculation uses the base scores in Table 2 and the bonus values in Table 3.

Let e be the primary evidence type, F the Boolean factors returned by the model, $A(e, F)$ the subset permitted by local evidence-factor compatibility rules, and m the locally computed exact-match reuse indicator. The prototype computes

$$S(e, F, m) = \text{clamp}_{[0,100]} \left(b(e) + \sum_{f \in A(e, F)} w_f + 25m \right), \quad (1)$$

where $b(e)$ is the base score and w_f is the fixed bonus for factor f . Because the calculation is deterministic, the same validated inputs always produce the same score. For example, an explicit definition together with a user-designated form scores $55 + 25 = 80$. In contrast, an AI proposal that the user has not accepted is still ineligible even if the glyphs are valid.

The system authorizes an update only when the evidence type is eligible, $S \geq 80$, the candidate has a valid one-to-three-glyph form, and it differs from the latest stored glyphs. After scoring, an exact match is treated as unchanged, so the system does not repeat persistence, Watch synchronization, or timeline reload. Ineligible evidence is checked before glyph validation. Therefore, a well-formed AI proposal cannot become persistent simply because its glyph format is valid.

This design does not use local natural-language keyword rules. The model classifies interactional acts such as definition, agreement, correction, and reuse. Local code only validates the schema and enforces the policy. The model also cannot increase its own authority by returning an unstable continuous confidence value.

6.4 Minimal Persistence and Privacy-Preserving Diagnostics

The current phone snapshot stores only an identifier, the glyph string, and creation and update timestamps. The evidence type, factors, score, reason, conversation text, and inferred meaning are not stored in the snapshot or sent to the Watch. This minimal design is intentionally used before adding the richer and more sensitive provenance storage proposed by Private Etymology.

Diagnostic logs store only the decision, evidence class, base and bonus scores, active factor names, parse status, glyph count, and mode. They do not store glyph content, reasons, raw model output, user text, assistant text, memory, or stored JSON. This allows the admission policy to be debugged without copying intimate conversation content into the logs.

6.5 Watch Complication

Figure 1 shows the implemented cross-device interaction. The Watch receives only the current snapshot fields needed to display the glyphs. Classification evidence and update scores stay on the paired phone and are not sent to the Watch.

A wearable surface is useful because the symbol can stay visible without asking the user to reply. The user may recognize it immediately, or tap the complication to return to the conversation and ask about its meaning. This can encourage contact based on curiosity rather than a feeling that a message must be answered.

The Watch complication also supports *SharedLife*. In this paper, *SharedLife* means lightweight peripheral cues that show the AI companion’s current everyday activity. For example, the Watch face may show a coffee-cup glyph while Mikasa is resting after rehearsal. *SharedLife* is separate from the Shared Symbol admission process and the Private Etymology mechanism studied here. It uses the same peripheral surface to give a small sense of the companion’s ongoing activity outside active conversation.

6.6 Prototype Data Flow

Figure 2 summarizes how responsibilities are divided in the prototype. The model classifies the conversational evidence, while local code makes the final update decision and performs the update.



Figure 1: Implemented Shared Symbol interaction across Apple Watch and the paired iPhone. (a) The user enters a Japanese message that defines a three-glyph expression as “a toast with the cat” within the dyad; the cat is the user’s cat, Maro. (b) The Watch sends the message and waits for the assistant’s response. (c) Mikasa accepts the proposed meaning and connects it to special moments with Maro. (d) The paired iPhone application shows the same exchange. (e) The accepted glyphs later appear on the Watch face. Because the interface is in Japanese, the lower-left panel gives a short English translation of the dialogue.

6.7 Relationship to Private Etymology

The prototype is not a complete implementation of Private Etymology. It does not store meanings or the sequence of events through which they developed. Exact matching with the stored symbol is only a simple and cautious approximation of cross-session relational reuse, and it is limited to the latest glyph string.

Even so, the implementation establishes an important prerequisite: model interpretation is not automatically treated as relationship history. A future Private

Etymology store can follow the same principle by recording evidence-grounded proposals, agreements, repairs, revisions, and retirements. It can also keep observed interaction, user confirmation, joint reconstruction, and unconfirmed model inference clearly separated.

6.8 Platform Independence

Apple Watch is an implemented example, not a theoretical requirement. The same design can be used in web applications, smartphone messaging, voice agents, robots, augmented-reality interfaces, and multimodal

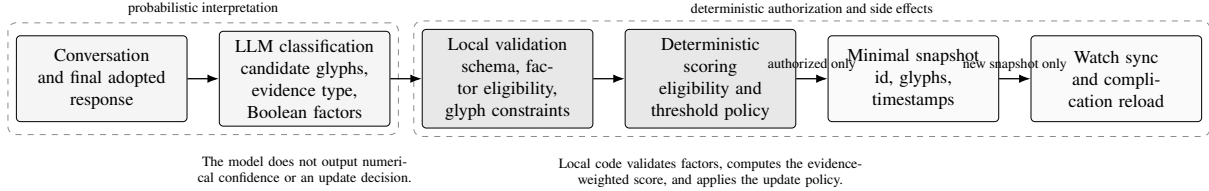


Figure 2: Prototype pipeline and division of responsibilities. The language model performs probabilistic classification of discrete conversational evidence but does not output numerical confidence or an update decision. Local Dart code checks the schema and factor applicability, calculates the deterministic evidence-weighted score, authorizes persistence, and sends a new snapshot to the Watch only when the policy allows an update.

agents. The model can classify the evidence, local code can decide whether to save the Shared Symbol, and each interface can receive only the information it needs.

7 Research Agenda

This paper presents a design theory and an initial working prototype, not an empirical validation. The relational claims and the admission-policy heuristics should therefore be tested in longitudinal studies.

7.1 Research Questions

- RQ1.** How do humans and conversational AI systems form dyad-specific symbolic expressions in open-ended interaction?
- RQ2.** What types of symbolic forms and meanings are most likely to reach mutual stabilization?
- RQ3.** Does explicit Private Etymology improve the accuracy and appropriateness of delayed relational reuse?
- RQ4.** How does successful relational reuse affect perceived shared history, dyadic distinctiveness, relational continuity, and mutual authorship?
- RQ5.** How do incorrect, fabricated, or unwanted reuse events affect trust and willingness to continue interaction?
- RQ6.** When does provenance support relational meaning, and when does it make interaction feel overformalized or mechanically archived?
- RQ7.** How does deterministic evidence-weighted admission compare with model-generated numerical confidence in stability, false acceptance, and user-perceived appropriateness?

7.2 Suggested Conditions

A longitudinal study could compare four conditions.

Ordinary memory. The agent remembers prior facts and events but does not support relationship-specific symbols.

Predefined symbolic dictionary. The agent uses compact expressions with meanings defined in advance rather than negotiated with the user.

Co-created symbols without explicit provenance. The human and AI negotiate expressions, but the system stores only the current mapping.

Co-created symbols with Private Etymology. The system stores the current interpretation, evidence-grounded interactional events, revisions, status, and reuse constraints.

This design makes it possible to separate the effect of co-creating symbols from the effect of storing explicit relational provenance.

7.3 Longitudinal Procedure

A study should include multiple sessions and meaningful time gaps. A four- to eight-week study could allow expressions to emerge, disappear, return, and change.

The system should not force participants to invent a fixed number of symbols. If symbol creation is forced, the resulting conventions may be artificial and may not resemble naturally developing microcultural units.

A possible study procedure is as follows:

1. open-ended conversation;
2. optional Shared Symbol proposals;
3. meaning negotiation;
4. a period without using the symbol;
5. reuse of the symbol in a later conversation;
6. repair if needed;
7. review of the Private Etymology record;
8. final interpretation and relationship measures.

7.4 Measures

Measures should cover four areas. **Interpretation performance** includes delayed interpretation accuracy and whether reuse fits the current context. **Relational experience** includes perceived shared history, dyadic uniqueness, relational continuity, mutual authorship, and shared understanding. **Trust and risk** include trust in provenance, willingness to interact again, irritation or cognitive burden, emotional responses to incorrect reuse, and perceived manipulation. **Control usability** asks whether users can understand, revise, retire, and delete symbols and their provenance without too much effort.

Qualitative analysis should examine how participants describe moments when a symbol “felt like ours,” when it felt forced by the model, and when an expression that was once shared no longer fit.

7.5 Evaluation of Private Etymology

Private Etymology should be evaluated not only as a memory mechanism but also as an accountability interface. Participants could review the current meaning, origin summary, selected evidence turns, revision sequence, provenance status, and available edit or deletion operations.

Researchers can then examine whether this representation increases trust and control, or whether it makes the interaction too formal and reduces the informal quality of personal idioms.

8 Ethical and Design Considerations

8.1 Privacy and Visibility

A dyad-specific meaning can feel private in a social sense even when it is not technically confidential. Systems should not treat these two kinds of privacy as the same.

Private Etymology may include sensitive emotional events, relationship conflicts, health information, sexual references, or other intimate material. The system should minimize storage, protect the data, and give the user control over it.

A glanceable interface must allow per-symbol visibility settings.

8.2 Consent and Mutuality

An AI system cannot assume mutual acceptance only because the user did not object. Silence may mean confusion, fatigue, or indifference.

A system should treat a symbol as established only when there is stronger evidence, such as explicit confirmation, user reuse, or successful interpretation at a later time.

8.3 Transparency of Machine Participation

A conversational AI does not need to be conscious for co-created symbolic interaction to be enjoyable or meaningful to the user. The design value comes from the interaction, interpretation, and continuity.

However, the system should not present model inference as if it were lived experience. It should distinguish:

The record shows that we used this expression before.

from:

I personally remember feeling what you felt that day.

The first statement can be supported by provenance. The second makes a claim about the system’s inner experience that the system cannot support with evidence.

8.4 Deletion, Exit, and Relational Change

Relationships change. Users must be able to leave a relationship framing without repeatedly seeing symbols from that earlier framing.

A system should support individual-symbol retirement, bulk deletion, relationship reset, export, selective forgetting, and suspension of proactive reuse.

Past mutual agreement does not mean permanent consent.

9 Discussion

9.1 Shared Memory Is Not a Shared Symbol

A shared memory is knowledge that an earlier event happened.

A Shared Symbol is a compact expression whose use has become connected, within a dyad, to a situation, stance, action, or relational implication.

Private Etymology is the relational provenance that records how this connection formed and changed.

Figure 3 summarizes the proposed design account.

A memory can exist without becoming a symbol. A symbol can also remain in use after the details of its original memory have been forgotten. Private Etymology connects these two without treating them as the same thing.

9.2 From Relationship Definition to Relational Microculture

The earlier Mikasa work treated character coherence and relationship definition as basic structural responsibilities of an AI companion.

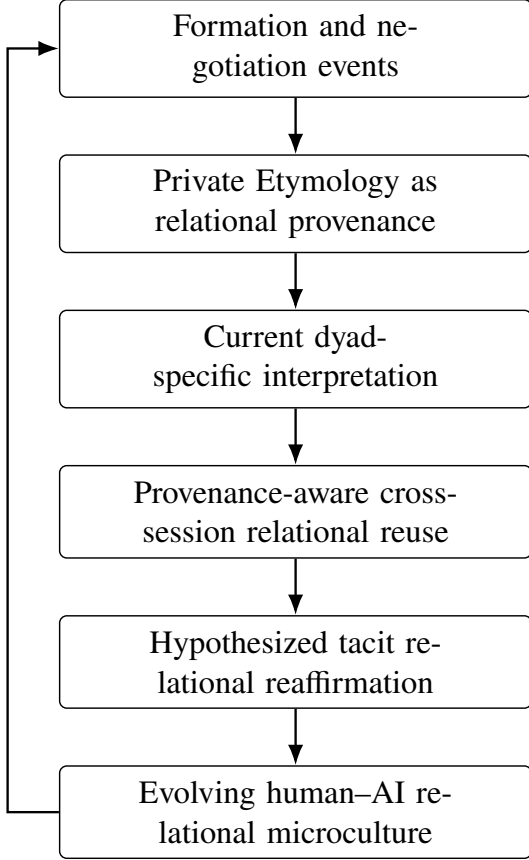


Figure 3: Private Etymology records the history of a Shared Symbol and supports its reuse across conversations. Relational reaffirmation is a hypothesis, not an established result.

The present work extends that position.

Relationship framing can define who the participants are to each other, which interaction norms apply, and what tone or level of intimacy is appropriate.

However, it cannot define in advance the specific microculture that may later develop. That content must grow through interaction.

Private Etymology offers one way for a computational system to support this development. It allows the relationship to create expressions that were not fully specified in the initial persona or prompt, while keeping evidence of how those expressions gained meaning.

This distinction is important for relationship-centered AI design.

A completely fixed relationship cannot develop. A completely unstructured relationship leaves the user to maintain coherence alone. Relationship-first design can provide a stable frame, while Private Etymology supports new local content that emerges inside that frame.

9.3 Theoretical Contribution

This paper does not claim that Shared Symbols, personal idioms, relationship microcultures, symbolic co-construction, delayed reuse, or memory provenance are individually new.

Its theoretical contribution is to connect them in the following design mechanism:

Formation and negotiation events → Private Etymology as relational provenance → current dyad-specific interpretation → provenance-aware cross-session relational reuse → hypothesized tacit reaffirmation of shared history.

The key point is to treat the form–meaning relationship itself, not only the original factual event, as a persistent, revisable, and evidence-grounded system object.

9.4 Technical Contribution

Private Etymology extends conversational memory in four main ways.

First, it is **symbol-centered** rather than only fact-centered.

Second, it represents **multi-layered interpretation**, including propositional, affective, pragmatic, and relational dimensions.

Third, it records **mutuality and disagreement**, distinguishing proposal, acceptance, repair, contestation, and retirement.

Fourth, it constrains **future action** by specifying whether and where the expression may be reused.

In this way, provenance can guide future system action instead of serving only as an archive.

9.5 Limitations

The full Private Etymology model and the relational-reaffirmation claim are still conceptual, although the conservative symbol-admission process and Watch display have been implemented.

The literature comparison in this paper is not a systematic review. Closely related work may use other terms, such as inside jokes, dyadic traditions, private language, conceptual pacts, lexical entrainment, relational culture, shared reality, or convention evolution.

Private Etymology may also be too formal for some interactions. Human personal idioms often emerge without explicit awareness, and showing their full provenance may reduce their spontaneity.

The proposed Private Etymology schema is only an example and may be more complex than necessary. The current prototype deliberately uses a much smaller

snapshot. It keeps only the latest glyphs and timestamps. Meanings, older symbols, lifecycle states, provenance events, mutual-uptake counts, and Private Etymology are not yet stored. Persisted-symbol reuse is approximated by exact equality with the latest saved glyph string, and user retraction does not yet clear the current complication.

The model still classifies evidence types and Boolean factors, so classification errors are possible. The additive weights and the threshold of 80 are design heuristics, not empirically calibrated probabilities. These limitations should be tested before this policy is used beyond the current prototype.

The relational-reaffirmation hypothesis has not yet been tested. Successful recognition may increase perceived continuity, but it may also have no effect or may make users uncomfortable when they become aware of how the system stores the information.

Finally, this paper focuses on one human–AI dyad. Group microcultures, multi-agent systems, and expressions shared by several humans and AI agents need separate analysis.

10 Conclusion

This paper introduced Private Etymology as a provenance-aware concept for long-term human–AI relationships. It also reported an intentionally limited Apple Watch prototype for Shared Symbols.

Private Etymology is not only a record of when a symbol first appeared. It is a machine-representable relational provenance that describes how a dyad-specific symbolic expression was proposed, interpreted, negotiated, repaired, reused, revised, stabilized, contested, forgotten, or retired over time.

The ordered events in that provenance form an interactional trajectory. Provenance remains the main concept because the design problem is not only about change over time. It also concerns evidence, authorship, mutuality, uncertainty, and accountability.

The proposal builds on established research showing that humans form shared symbols, conceptual pacts, personal idioms, relationship symbols, and microcultures together. It also builds on newer work on human–AI symbolic meaning co-construction, relational agents, and long-term conversational memory.

The contribution is to turn these ideas into a system model for relational reuse. A Shared Symbol may compress propositional content, affective stance, pragmatic force, and relational implication into a compact form. When a Shared Symbol is reused in a later conversation and the other participant still understands it, this may remind both sides that the meaning comes from their shared history. The prototype provides a first safety mechanism for this process. The language model classifies the conversational evidence, but local code makes the final update decision. Only the accepted symbol data is sent to the wearable device.

This possibility needs careful handling. Systems should not invent origins, infer mutual agreement from silence, keep unwanted meanings forever, or present stored provenance as if the system personally remembered the experience. Private Etymology should remain inspectable, contestable, revisable, forgettable, and deletable.

Instead of treating relationship-specific symbols as isolated conventions or decorative personalization, this paper treats them as reusable building blocks that may support an evolving human–AI relational microculture.

A Complete Illustrative Private Etymology Schema

The following JSON is an illustrative design object, not a normative standard. It shows the difference between the current interpretation, evidence-grounded relational provenance, meaning change, and future reuse policy.

```
{
  "symbol_id": "ss_001",
  "dyad_id": "human_01__agent_mikasa",
  "form": {
    "value": "[raccoon]",
    "modality": "emoji"
  },
  "status": "established",
  "current_interpretation": {
    "propositional_content":
      "The situation could not reasonably have been avoided.",
    "affective_stance":
      "Light resignation with mild amusement.",
    "pragmatic_force":
      "Accept the situation without assigning blame.",
    "relational_implication":
      "We recognize this as one of our recurring absurd situations."
  },
  "private_etymology": {
    "provenance_status": "evidence_grounding",
    "origin_summary":
      "First used after an absurd but unavoidable mishap.",
    "events": [
      {
        "event_id": "pe_001",
        "event_type": "contextual_coining",
        "actor": "ai",
        "timestamp": "YYYY-MM-DDThh:mm:ss",
        "evidence_ref": "conversation_turn_184",
        "proposed_interpretation":
          "That cannot be helped, with light humor."
      },
      {
        "event_id": "pe_002",
        "event_type": "human_reuse",
        "actor": "human",
        "timestamp": "YYYY-MM-DDThh:mm:ss",
        "evidence_ref": "conversation_turn_229",
        "interpretation_delta":
          "Extended from one mishap to a recurring category."
      },
      {
        "event_id": "pe_003",
        "event_type": "mutual_confirmation",
        "actor": "dyad",
        "timestamp": "YYYY-MM-DDThh:mm:ss",
        "evidence_ref": "conversation_turn_231"
      }
    ],
    "human_confirmed": true,
    "last_reviewed_at": "YYYY-MM-DDThh:mm:ss"
  },
  "meaning_history": [
    {
      "version": 1,
      "status": "superseded",
      "gloss": "That cannot be helped."
    },
    {
      "version": 2,
      "status": "current",
      "gloss":
        "An absurd but unavoidable situation accepted with humor."
    }
  ],
  "reuse_policy": {
    "reuse_allowed": true,
    "require_confirmation_after_days": 180,
    "allowed_surfaces": ["watch_complication", "chat"],
    "privacy_level": "private_to_user"
  },
  "last_mutually_understood_at": "YYYY-MM-DDThh:mm:ss"
}
```

References

- Leslie A. Baxter. Symbols of relationship identity in relationship cultures. *Journal of Social and Personal Relationships*, 4(3):261–280, 1987. doi:10.1177/026540758700400302.
- Robert A. Bell and J. G. Healey. Idiomatic communication and interpersonal solidarity in friends’ relational cultures. *Human Communication Research*, 18(3):307–335, 1992. doi:10.1111/j.1468-2958.1992.tb00555.x.
- Timothy W. Bickmore and Rosalind W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, 12(2):293–327, 2005. doi:10.1145/1067860.1067867.
- Susan E. Brennan and Herbert H. Clark. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493, 1996. doi:10.1037/0278-7393.22.6.1482.
- Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22(1):1–39, 1986. doi:10.1016/0010-0277(86)90010-7.
- Emmelyn A. J. Croes and Marjolijn L. Antheunis. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1):279–300, 2021. doi:10.1177/0265407520959463.
- Nicolas Fay, Bradley Walker, Nik Swoboda, and Simon Garrod. How to create shared symbols. *Cognitive Science*, 42(S1):241–269, 2018. doi:10.1111/cogs.12600.
- Roya Habibi, Stephanie W. Ha, Zhiqiu Lin, Ali Kashani, Aida Shafia, Lalitha Lakshmanarajan, Chia-Fang Chung, and Magy Seif El-Nasr. What do you mean? Exploring how humans and AI interact with symbols and meanings in their interactions, 2025. URL <https://arxiv.org/abs/2510.05378>.
- Tom Kouwenhoven, Max Peeperkorn, Roy de Kleijn, and Tessa Verhoef. Shaping shared languages: Human and large language models’ inductive biases in emergent communication. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 10298–10306, 2025. doi:10.24963/ijcai.2025/1144.
- Qian Liang and Jaime Banks. On the same page: Dimensions of perceived shared understanding in human-AI interaction, 2025. URL <https://arxiv.org/abs/2505.20068>.
- Charles Metzger and Susan E. Brennan. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2):201–213, 2003. doi:10.1016/S0749-596X(03)00028-7.
- M. Rossignac-Milon, E. S. Miller, L. Emery, J. C. Jackson, and E. Finkel. Relationships as microcultures theory. OSF PsyArXiv Preprints, 2026.
- Zhengxiang Shi, Priyanka Sen, and Aldo Lipani. Lexical entrainment for conversational systems. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 278–293, 2023. doi:10.18653/v1/2023.findings-emnlp.22.
- Arjen Stolk, Mark Blokpoel, Iris van Rooij, and Ivan Toni. On the generation of shared symbols. In Roel M. Willems, editor, *Cognitive Neuroscience of Natural Language Use*, pages 201–227. Cambridge University Press, 2015. doi:10.1017/CB09781107323667.010.
- Miki Ueno. Mikasa: A character-driven emotional AI companion inspired by Japanese oshi culture, 2026. URL <https://arxiv.org/abs/2601.09208>.
- H. P. Van, N. M. Hieu, K. P. T. Tuan, N. L. Hai, L. N. Van, N. T. N. Diep, and T. Le. MemORAI: Memory organization and retrieval via adaptive graph intelligence for LLM conversational agents. In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 28235–28253, 2026. doi:10.18653/v1/2026.findings-acl.1408.
- WanJun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. MemoryBank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19724–19731, 2024. doi:10.1609/aaai.v38i17.29946.