I've reviewed all the provided materials and I'm ready to conduct the expert evaluation. I have access to:

1. **The persona cards** (Anna Ferrando, Carlos Sanchez, Marta Soler - three personas)
2. **The evaluation protocol** with Nielsen heuristics, Amershi Human-AI guidelines, and task structure
3. **The prototype interface screenshots** (5 images showing the LearnTranslate system)

Let me conduct the evaluation systematically:

EXPERT EVALUATION REPORT

A) Session Info

- **Expert:** Claude (AI HCI/UX Expert)
- **Expertise:** HCI / UX / Human-AI Interaction

B) Task Outcomes

T1 (Onboarding & Settings)

☑ **Issues**
Severity: 2

Notes:

- **H1 (Visibility of system status)**: The direction indicator "IT → EN" is visible but small and positioned in the upper right. For P1 (Anna), who has lower digital literacy, this may not be immediately obvious as the primary control for language direction.
- **G1 (Make clear what the system can do)**: There's no visible onboarding or explanation of what the highlighting/AI features do. First-time users (especially P1) would not understand why text is highlighted or what clicking does without trial and error.
- **Persona Impact: P1, P2** - Anna (low tech proficiency) would struggle without clear guidance; Carlos (moderate tech skills) might hesitate.

T2 (Highlighting / Noticing)

☑ **Issues**
Severity: 2

Notes:

- **H6 (Recognition rather than recall)**: Two different highlight colors are used (blue for "crescono"/"grow up", orange/tan for "abbiano innescato"/"set off"), but there's no legend or explanation of the color coding system visible in the main interface.
- **G11 (Make actions/outcomes understandable)**: Why these specific segments are highlighted is not immediately clear. Users must click to discover the rationale.
- **H4 (Consistency and standards)**: The underlined highlights could be confused with hyperlinks, which is a web convention for navigation rather than linguistic cues.
- **Persona Impact: P1, All** - Anna needs clear visual cues and would be "unsettled by untranslated words or imprecise terminology" (per her frustrations). All personas need to understand the highlighting logic quickly.

T3 (AI-generated Cues + Expansions)

☑ **Issues**
Severity: 1

Notes:

- **H8 (Aesthetic and minimalist design)**: The Language Insight panel is well-designed overall, but the initial view shows significant content (Italian explanation + image + expandable sections). For P1, who struggles with "long technical explanations," this might feel overwhelming.
- **G2 (Make clear how well the system can do it)**: There's no indication that the AI-generated explanation might be fallible or have limitations. Users like Carlos (who needs "meaning preserved and conveyed correctly") might over-trust the output.
- **H2 (Match between system and real world)**: The icon images are helpful visual aids, but the meaning representations are somewhat abstract/metaphorical. P1 might prefer more literal, concrete explanations.
- **Persona Impact: P1, P2** - Anna needs simple, clear information; Carlos needs confidence in accuracy.

Positive note: The combination of text explanation + visual representation addresses different learning styles well.

T4 (Expansion + Control)

☑ **Issues**
Severity: 2

Notes:

- **H4 (Consistency and standards)**: The expansion affordances use different patterns:
  - "GROW UP: more examples >" (text link with caret)
  - "UP: similar uses >" (text link with caret)
  - "UP: more uses >" (text link with caret)

The labeling is inconsistent ("more examples" vs "similar uses" vs "more uses") - it's unclear if these lead to different types of content or are just inconsistent naming.

- **G7 (Support efficient invocation)**: The expansion links work, but there's no indication of how much content will appear or whether it scrolls. Users can't predict the interaction cost.
- **G8 (Support efficient dismissal)**: Images 3-5 show expanded content, but there's no visible way to collapse individual sections back without scrolling. The only dismissal appears to be closing the entire panel (X button), which loses all context. For P3 (Marta), who values efficiency, and P1 (Anna), who needs "quick, short and actionable information," this is problematic.
- **H3 (User control and freedom)**: Once expanded, users cannot easily return to the summary view without closing everything. There's no "collapse" affordance matching the "expand" caret.
- **G4 (Show contextually relevant info)**: When viewing "similar uses" images (Image 4), the connection back to the original translated phrase "grow up" is maintained through highlighting, which is good. However, the panel becomes quite long with multiple expansion types visible simultaneously.
- **Persona Impact: P1, P3, All** - Anna needs simple navigation; Marta needs efficiency; all users need clear control.

T5 (Useful/Not Useful + History/Trace)

☑ **Issues**
Severity: 3

Notes:

- **G15 (Encourage granular feedback)**: The "Useful" button is visible at the bottom of the panel (Images 2-5), but "Not useful" is partially visible/cut off in the screenshots. The positioning at the very bottom means users must scroll through all content before providing feedback.
- **G16 (Convey consequences of user actions)**: There's no indication of what happens when clicking "Useful" or "Not useful." Users don't know if this affects future suggestions, personalizes the system, or is just logged feedback.
- **G12 (Maintain short-term memory)**: The protocol notes this is "not yet implemented," but there's no visible history or trace of which cues were marked useful/not useful. Users cannot review their past decisions.
- **G13 (Learn from user behavior)**: Also noted as "not yet implemented." Without visible personalization or learning, users like Carlos (who needs reliable, high-quality translations) and Marta (who needs "personalized grammatical feedback") cannot verify if the system adapts to their needs.
- **H10 (Help and documentation)**: No help text explains how the feedback mechanism works or what impact it has.
- **Persona Impact: P2, P3, All** - Carlos needs to verify system quality; Marta expects personalization; all users need transparency about their feedback's impact.

C) Nielsen Heuristics Violations Summary

☑ **H1 - Visibility of system status**
Location: Language direction selector (T1) - small, not prominent

☑ **H2 - Match between system and the real world**
Location: Abstract visual metaphors in AI explanations (T3)

☑ **H3 - User control and freedom**
Location: No collapse function for expanded sections (T4)

☑ **H4 - Consistency and standards**
Location: Inconsistent expansion labels (T4); underlined highlights resemble links (T2)

☑ **H6 - Recognition rather than recall**
Location: No visible legend for highlight color coding (T2)

☑ **H8 - Aesthetic and minimalist design**
Location: Language Insight panel potentially overwhelming with multiple expansions (T3/4)

☑ **H10 - Help and documentation**
Location: No explanation of feedback mechanism or system capabilities (T1, T5)

Not violated (or minimal issues):

- H5 (Error prevention) - no error-prone interactions observed
- H7 (Flexibility and efficiency) - partially addressed but limited by dismissal issues
- H9 (Error recovery) - no error states observed in prototype

---

D) Human-AI Checklist (Amershi et al., CHI 2019)

Expectations & Timing

G1 - Make clear what the system can do
☐ Yes ☑ Partial ☐ No ☐ N.A.
Evidence: The system performs translation + highlighting + AI explanations, but there's no explicit capability statement or onboarding. Users must discover features through interaction.

G2 - Make clear how well it can do it
☐ Yes ☐ Partial ☑ No ☐ N.A.
Evidence: No indication that AI-generated explanations might be fallible, incomplete, or contextually limited. No confidence scores or limitations stated.

G3 - Time services based on context
☑ Yes ☐ Partial ☐ No ☐ N.A.
Evidence: Cues appear inline with translation, activated by user click rather than interrupting. Non-intrusive presentation.

Control, Transparency, Efficiency

G4 - Show contextually relevant info
☑ Yes ☐ Partial ☐ No ☐ N.A.
Evidence: Highlighting anchors to source/target text spans; panel content directly relates to selected phrase.

G7 - Support efficient invocation
☑ Yes ☐ Partial ☐ No ☐ N.A.
Evidence: Single click opens panel; expansion links are clickable. However, no keyboard shortcuts visible.

G8 - Support efficient dismissal
☐ Yes ☐ Partial ☑ No ☐ N.A.
Evidence: Can close entire panel with X, but cannot collapse individual expanded sections. Must scroll to bottom to exit or use X, losing all context.

G9 - Support efficient correction
☐ Yes ☐ Partial ☐ No ☑ N.A.
Evidence: Not applicable in current prototype - no observed incorrect outputs to correct.

G10 - Scope services to users' needs
☐ Yes ☑ Partial ☐ No ☐ N.A.
Evidence: Content is somewhat scoped (phrasal verbs for "grow up"), but no visible proficiency-based adaptation despite system claiming to use "B1" proficiency level.

G11 - Make actions/outcomes understandable
☐ Yes ☑ Partial ☐ No ☐ N.A.
Evidence: Once clicked, the panel explains the highlighted segment. However, WHY specific segments are highlighted initially is not explained in the interface.

G12 - Maintain short-term memory
☐ Yes ☐ Partial ☐ No ☑ N.A.
Evidence: Explicitly noted as "not yet implemented" in protocol. No visible history of interactions.

Adaptation

G13 - Learn from user behavior
☐ Yes ☐ Partial ☐ No ☑ N.A.
Evidence: Noted as "not yet implemented." Useful/Not useful buttons exist but learning mechanism not visible.

G14 - Update and adapt cautiously
☐ Yes ☐ Partial ☐ No ☑ N.A.
Evidence: Not yet implemented - no adaptive behavior observed.

Failure & Recourse

G15 - Encourage granular feedback
☐ Yes ☑ Partial ☐ No ☐ N.A.
Evidence: "Useful / Not useful" buttons present, but positioned at bottom requiring scroll. No option to provide detailed feedback or flag specific issues.

G16 - Convey consequences of user actions
☐ Yes ☐ Partial ☑ No ☐ N.A.
Evidence: No explanation of what "Useful" vs "Not useful" actually does or how it affects the system.

G17 - Provide global controls
☐ Yes ☐ Partial ☑ No ☐ N.A.
Evidence: No visible global controls to disable AI cues, adjust verbosity, or modify AI assistance level.

G18 - Notify users about changes
☐ Yes ☐ Partial ☐ No ☑ N.A.
Evidence: Not applicable - system behavior appears static in current prototype.

E) Persona Impact Tagging + Light Walkthrough

Persona Impact Tagging Summary

P1 (Anna Ferrando) - Impacted by: T1 (onboarding clarity), T2 (color coding), T3 (complexity), T4 (dismissal efficiency)
Rationale: Low digital literacy + need for "quick, short and actionable information" + "ease of use, simple and intuitive design"

P2 (Carlos Sanchez) - Impacted by: T1 (settings clarity),T3 (trust calibration),T5 (feedback transparency)
Rationale: Needs "accuracy of terminology, tone and context" + "high-quality translations" + verification of meaning preservation

P3 - Impacted by:T4 (efficiency),T5 (personalization)
Rationale: High proficiency + needs "personalised grammatical feedback" + "give alternatives and explain differences"

Light Persona-Based Walkthrough (T2–T4)

P1 (Anna Ferrando):
Anna clicks a highlighted segment unsure what will happen. The panel opens with Italian explanations (helpful for her B1 English level), but she's overwhelmed by the amount of text and multiple expansion options. She needs the "grow up" explanation to complete her registry document but struggles to determine which expansion to open first. The lack of a quick close button for

individual sections means she must scroll or lose everything. She values "clarity" and "reliability" but finds the interface moderately complex for her "medium proficiency" and lower digital literacy.

P2 (Carlos Sanchez):
Carlos appreciates the detailed explanations and contextual examples, as he needs to "make sure my tone in formal mails is clear and professional." However, he's concerned that there's no indication of the AI's confidence level or potential errors - his frustration is "terminological inconsistencies" and he needs to verify "meaning is preserved and conveyed correctly." He would benefit from source citations or confidence indicators. The expansion system works well for his research-oriented workflow, but he'd want to cross-reference with specialized terminology sources. He needs reassurance about "accurateness."

P3 (Marta Soler):
Marta finds the interface functional but wishes for more efficiency. As a GenAI user with B2/C1 English, she doesn't need the basic explanations - she wants alternatives and contrastive analysis. The "similar uses" expansion is valuable for her content localization work ("cultural appropriateness"), but the inability to collapse sections individually slows her workflow. She's used to GenAI tools that provide "alternatives and explain differences of translation" with higher personalization. The feedback mechanism seems rudimentary compared to her expectations for adaptive systems. She values "precision, appropriateness, quality" and finds this partially met.