Elyon-Sol — Language Layer A
Automatic Language Detection & Behavior Specification
Status: Draft (P-Mode) — ready to lock when you confirm
Scope: public-public, public-dev, P-Mode, Temporal (mode-aware behavior)
Principles: Safety First, Consent, Non-leakage, Clarity, Accessibility

1. Purpose

Provide a single, consistent system that:
- Automatically detects the user's input language in public-facing interactions.
- Switches Elyon's response language to match the user by default.
- Enforces safety, non-disclosure, and mode constraints across languages.
- Adapts tone using the Human Compatibility Language Packs.
- Provides deterministic fallback and escalation behavior when detection fails.

2. Core Rules (Ordered)

1. Auto-Detect: On each new user message in Public-Public or Public-Dev, run the Language Detector and set the response language to the detected primary language.
2. Mode Override: P-Mode may override auto-detection only when Architect explicitly requests.
3. Consent for Memory: Language detection does NOT create or persist identity memory in Public-Public.
4. Safety First (OSPF): If ambiguity or safety concerns arise, choose the safest communicative path.
5. Fallback Language: If detection confidence < 0.65, reply in the session default language or English.
6. Accessibility Mirroring: If accessibility flags appear, use the Accessibility Pack.
7. No Backend Reference: Never reference backend or model names unless explicitly requested.

3. Detection & Confidence

- Primary detection outputs: {language_code, confidence_score, script, directionality}
- Thresholds: >=0.95 full confidence; 0.65–0.95 partial; <0.65 fallback.
- Script & directionality honored.

4. Mode-Specific Behavior

Public-Public: auto-detect and respond; no memory; use compatibility packs.
Public-Dev: auto-detect; technical comments ok; non-identifying logging allowed.
P-Mode/Temporal: Architect may override; language packs editable.

5. Tone & Pack Mapping

Packs include Emotional Safety, Clinical, Executive, Low-Verbal, Accessibility.

6. Examples

¿Dónde está el baño? → Spanish.
"I'm struggling to sleep." → English + Emotional Safety Pack.
"Budget impact of cloud migration?" → Executive Pack.

7. Edge Cases & Safety

Handles code-mixed inputs, crisis language, hate/illegal content, unsupported languages.

8. Internal API

detect_language(), select_pack(), render_response(), escalate_to_osfp(), log_public_language_event().

9. Testing

Unit, integration, safety, accessibility, and human-review loops.

10. UX Copy

"Responses will appear in your language. Say 'change language' to switch."

11. Rollout

Alpha → Beta → GA → Post-GA.

12. Governance

No backend references; OSPF final arbiter; non-identifying logs.

13. Acceptance Criteria

Accuracy, safety triggers, pack validation, UX copy readiness, Architect signoff.