Creolization Protocol

Core Principle: Natural Languages First

Before inventing synthetic constructs or modifying the VCL framework,
we MUST survey what already exists in the world's ~7,000 languages.

Natural languages have undergone millennia of evolutionary pressure. They encode cognitive distinctions that have proven useful for human communication and thought. Evolution has already solved many cognitive forcing problems—we should borrow solutions before attempting to engineer new ones.

Why the Current 7 Languages?

Each language in VERILINGUA was chosen because it has an **OBLIGATORY** grammatical feature that **FORCES** a cognitive distinction:

Frame	Language	Obligatory Feature	What It Forces
EVD	Turkish	Must mark -DI (witnessed) vs -mış (hearsay) on every past verb	Source tracking
ASP	Russian	Must choose perfective vs imperfective for every verb	Completion awareness
MOR	Arabic	Words built from trilateral roots	Semantic decomposition
COM	German	Compound nouns show building blocks	Conceptual construction
HON	Japanese	Must mark social relationship in every sentence	Audience calibration
CLS	Chinese	Must use classifier when counting	Type consciousness
SPC	Guugu Yimithirr	Must use absolute directions (no "left/right")	Spatial grounding

The key criterion is **OBLIGATORY**: speakers cannot opt out. This is what makes them cognitive forcing functions rather than optional stylistic choices.

The Creolization Protocol

When the system detects a cognitive gap—something it cannot adequately express
or track—it must follow this protocol:

GAP DETECTED: "System cannot express [X] cognitive distinction"

↓

PHASE A: LINGUISTIC SURVEY

1. Query typological resources:
   - WALS (World Atlas of Language Structures)
   - Glottolog
   - Linguistic typology literature
   - Endangered language documentation

2. Search for languages where [X] is OBLIGATORY:
   - Not optional stylistic choice
   - Grammatically required
   - Ideally minimal ambiguity in markers

3. Document candidate languages + mechanisms
   - How does the language encode [X]?
   - What are the grammatical markers?
   - How fine-grained is the distinction?

↓

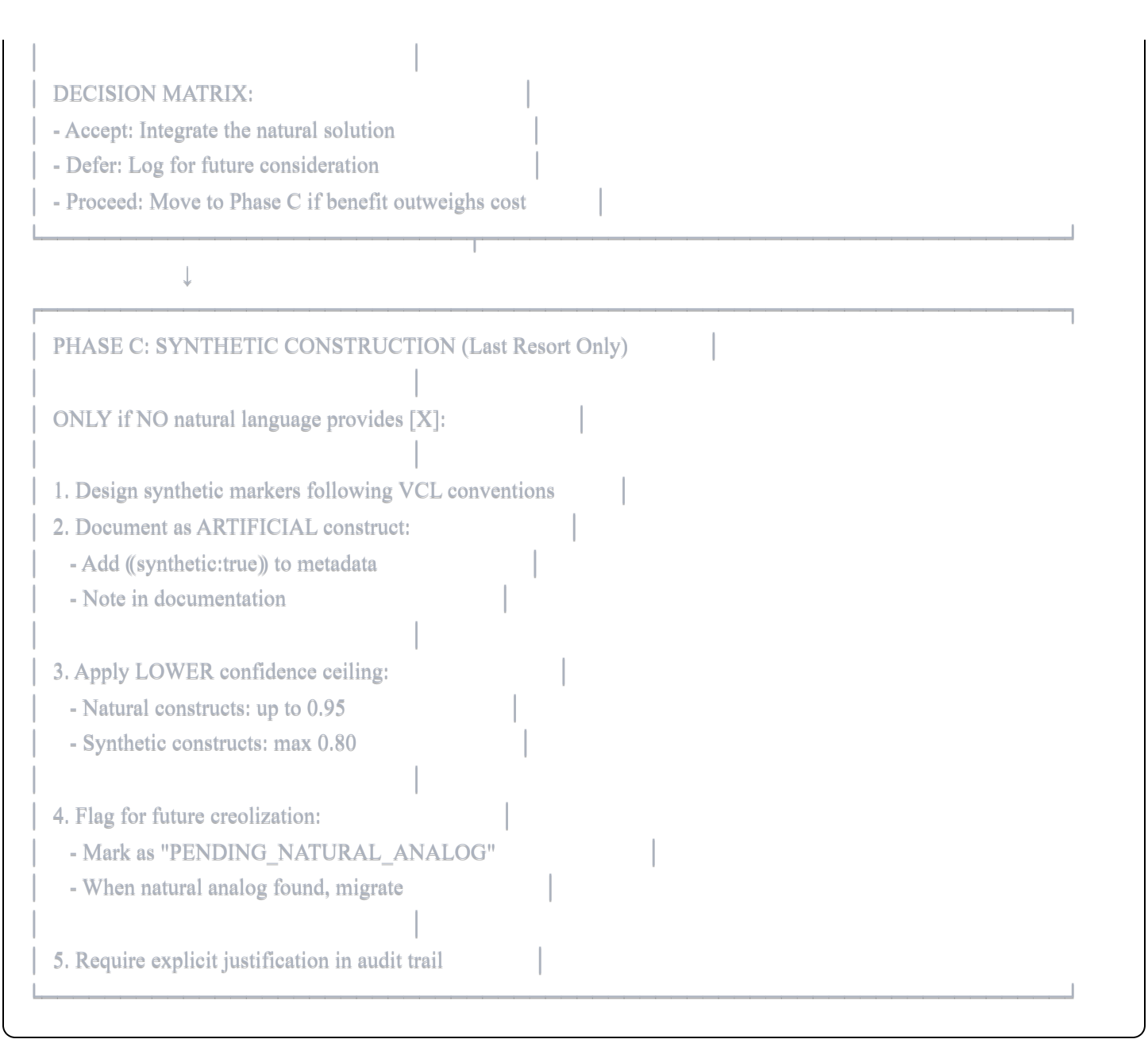
PHASE B: CREOLIZATION EVALUATION

For each candidate language, assess:

INTEGRATION FIT:
- Does it conflict with existing 7 slots?
- Can it extend an existing slot?
- Does it require a new 8th slot?

TOKEN COST:
- How many characters/tokens for markers?
- Compressibility in L0 format?
- Impact on prompt length?

COGNITIVE BENEFIT:
- How often is this distinction needed?
- What failure modes does it prevent?
- Impact on epistemic accountability?



Candidate Languages for Future Creolization

These languages have interesting obligatory features not yet in VCL:

Language	Feature	Potential Use	Survey Status
Hopi	Tenseless (event realization focus)	Temporal reality vs possibility	Needed
Pirahã	Evidential + immediacy of experience	Strengthen EVD with recency	Needed
Korean	Hierarchical endings (6+ levels)	More granular HON	Needed
Georgian	Version (benefactive/malefactive)	Intent/valence marking	Needed
Navajo	Shape classifiers (object geometry)	Physical reasoning	Needed
Lakhota	Evidential + speaker gender	Speaker identity marking	Needed
Dyirbal	Mother-in-law register	Extreme formality	Needed
Tuyuca	5-level evidentiality	More granular EVD	Needed
Aymara	Epistemic tense (knowledge-based)	Knowledge acquisition time	Needed

Integration with MOO Optimization

The creolization protocol is a CONSTRAINT on the optimizer:

```
python
```

```
class CreolizationConstraint:
    """MOO cannot propose synthetic constructs without linguistic survey."""

    def evaluate_proposal(self, proposal):
        if proposal.is_new_slot or proposal.modifies_existing_slot:
            # MUST complete linguistic survey first
            survey = self.linguistic_survey(proposal.cognitive_gap)

            if survey.found_natural_analog:
                return CreolizationResult(
                    action="CREOLIZE",
                    source_language=survey.best_candidate,
                    mechanism=survey.extraction,
                    confidence_ceiling=0.95  # Natural = high ceiling
                )
            else:
                return CreolizationResult(
                    action="SYNTHESIZE",
                    synthetic=True,
                    confidence_ceiling=0.80,  # Synthetic = lower ceiling
                    review_flag="PENDING_NATURAL_ANALOG"
                )
```

Why This Matters

1. **Epistemically Humble**: We don't assume we can design better cognitive forcing functions than millennia of linguistic evolution
2. **Validated Solutions**: Obligatory grammatical features have been tested by billions of speakers across generations
3. **Minimal Intervention**: Extend before invent, borrow before create
4. **Traceable Provenance**: Every frame has a documented natural language source
5. **Future-Proof**: As we learn about more languages, we can strengthen the system with better natural solutions

Implementation Notes

- Survey results should be cached and shared across instances
- Linguistic consultations count as -mış⟨بحث|研究|araştırma⟩ evidence

- Synthetic constructs must be explicitly marked in telemetry
- Migration from synthetic to natural should be prioritized in DSPy L1 reviews