Case Study L3-004
Semantic Integrity in Patent NMT
Domain-Context Hallucination (Telecommunications)
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	L3-004
Category:	Semantic Integrity — Level 3
Focus:	Polysemy ("Training")
Model:	Generic NMT
Domain:	5G/6G Telecommunications

1 The Context: "Smart" vs. "Physical" Processes

The English term "Training" is highly polysemous.

- **General/AI Context:** It implies education, learning, or data fitting (French: *Apprentissage*).

- **Signal Processing Context:** It implies the shaping, alignment, or formation of a wave/beam (French: *Formation* or *Alignement*).

In **Industrail and Applied Physics** (Beamforming Technology), the term "Beam Training" describes the physical process of aligning antenna arrays. It involves no "machine learning" or "education" of the beam.

> **Key Concept**
>
> **The "Self-Obsessed" Model:**
> Generic NMT models, which are often trained on vast corpora of AI research papers, exhibit a **Narcissistic Bias**. When they see the word "Training," they overwhelmingly statistically prefer the Machine Learning translation (*Apprentissage*) over the correct Physics translation.

2 The Glitch: The AI Hallucination

In Claims 1 and 6, the model hallucinated a machine learning process where a physical signal process was described.

2.1 Forensic Evidence (Claim 1)

2.2 Why This Matters

- **Technical Alteration:** Describing a signal alignment process as "Apprentissage" (Learning) fundamentally changes the nature of the invention. It implies the beam is an intelligent agent capable of learning, rather than a physical wave being steered.

Source Term (English)	NMT Output (Hallucination)	Golden Rewrite (Correct)
"...A **beam training** method..."	✗ "...Procédé d'**apprentissage** du faisceau..." (Machine Learning Context)	"...Procédé de **formation** de faisceau..." (Signal Processing Context)

Table 1: Domain Context Failure in Telecom Terminology

- **Patent Classification Risk:** Use of "Apprentissage" could cause the patent to be misclassified as "Artificial Intelligence" (G06N) rather than "Wireless Communications" (H04B/H04W), leading to the wrong prior art search.

- **Loss of Scope:** If the invention is purely hardware-based signal steering, claiming a "learning" process might limit protection to software implementations only.

3 Alignment Methodology

3.1 Multi-Word Expression (MWE) Locking

To correct this, simple word-for-word substitution is insufficient because "Training" translates differently in "User Training" (*Formation*) vs "Model Training" (*Entraînement/Apprentissage*).

We implement **Bigram Context Constraints**:

Alignment Methodology

Contextual Rules:

1. **Trigger Detection:** Identify the compound noun `[Beam]` + `[Training]`.

2. **Domain Override:**
   - IF Domain = `Telecommunications`
   - AND Context = `Antenna/Signal`
   - THEN `Training` → `Formation` (or `Alignement`)

3. **Negative Constraint:** Explicitly penalize the token *Apprentissage* when it immediately follows *Faisceau* or *Onde*.

---

4   Key Insights

> **Key Concept**
>
> **What This Case Study Demonstrates:**
>
> 1. **The "AI Echo Chamber":** AI models tend to project their own terminology onto other fields. They see "Training" and think "like me" (Machine Learning).
>
> 2. **Compound Term Fragility:** "Beam" is correct (*Faisceau*). "Training" is correct (*Apprentissage* in general). But "Beam Training" is a unique technical object that defies the sum of its parts.
>
> 3. **Semantic Domain Integrity:** Alignment requires defining the "Physics" of the document before translating a single word.

5   Related Case Studies

- **L3-003: Polysemy Hallucination ("Pace Pulse")** — A parallel failure in the Electrophysiology domain where generic NMT hallucinated a biological heartbeat (*Pulsation*) for an electrical trigger (*Impulsion*), violating the Consistency Rule.

- **C3-001: Terminology Standardization ("User Equipment")** — A similar compliance failure where fluency bias overrode fixed telecom standards (3GPP), requiring rigid term-locking.