Case Study L1-009
Semantic Precision in Patent NMT
Polysemy & Oscillation (The "Tiny Bus Driver")
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	L1-009
Category:	Semantic Precision — Level 1
Focus:	Mechanical vs. Human Agents
Model:	Generic NMT
Domain:	MedTech / Mechanical Engineering

1   The Context: Mechanical Actuation

In mechanical engineering, a **"Driver"** is a component (gear, cam, motor) that transmits motion or force to another part. In French, the correct technical term for this mechanical agent is **"Entraîneur"** or **"Organe moteur"**.

> **Key Concept**
>
> **The Polysemy Trap:**
> The word "Driver" is highly context-dependent:
>
> - *Bus Driver* → **Conducteur** (Human operator).
> - *Copper Driver* → **Conducteur** (Electrical wire).
> - *Software Driver* → **Pilote** (IT).
> - *Gear Driver* → **Entraîneur** (Mechanics).
>
> Generic NMT models often default to the highest frequency meaning ("Conducteur") regardless of the mechanical context.

2   The Glitch: The "Tiny Bus Driver"

In Claim 1 of a medical inhaler patent, the generic model introduced a human operator into a microscopic device. Even worse, it exhibited "Goldfish Memory" by switching terms in Claim 3.

2.1   Forensic Evidence (Claim 1 vs. Claim 3)

2.2   Why This Matters

- **Physical Impossibility:** Translating "Driver" as "Conducteur" implies the inhaler contains either a human being or an electrical wire. Since the device is purely mechanical (DPI),

Source Term (English)	NMT Output (Hallucination)	Golden Rewrite (Correct)
Claim 1: "...a **driver** (2)..."	✕ "...un **conducteur** (2)..." (Meaning: Bus Driver / Wire)	"...un **entraîneur** (2)..." (Meaning: Drive Gear)
Claim 3: "...the **driver** engages..."	"...l'**entraîneur** engage..." (Inconsistent with Claim 1)	"...l'**entraîneur** engage..." (Consistent)

Table 1: Polysemy & Consistency Failure in MedTech

"Conducteur" suggests an electrical hazard that doesn't exist.

- **Consistency Violation:** The oscillation between "Conducteur" (Claim 1) and "Entraîneur" (Claim 3) violates the fundamental patent rule: *Same Term = Same Feature*.

- **Indefiniteness:** A court could rule that "Conducteur" and "Entraîneur" refer to two different components, rendering the claims ambiguous.

3 Alignment Methodology

3.1 Variable-Based Domain Locking

We treat polysemous nouns as **Context-Dependent Variables**.

Alignment Methodology

The "Driver" Disambiguation Protocol:

1. **Domain Detection:** Identify context markers: `[Inhaler]`, `[Gear]`, `[Rotation]`, `[Actuator]`.

2. **Vector Locking:**

   - IF Domain = `MECHANICAL`:
   - **Ban:** *Conducteur* (Human/Wire), *Chauffeur* (Driver), *Pilote* (Software).
   - **Force:** *Entraîneur* or *Organe d'entraînement*.

3. **Global Consistency:** The system scans the entire document. If "Entraîneur" is chosen in Claim 3 (high confidence), it back-propagates to correct "Conducteur" in Claim 1 (low confidence).

---

4   Key Insights

> **Key Concept**
>
> **What This Case Study Demonstrates:**
>
> 1. **The "Goldfish" Effect:** Generic models often translate sentence-by-sentence, forgetting their own terminology choices from previous paragraphs.
>
> 2. **Agent Confusion:** The AI struggled to distinguish between an "Agent of Motion" (Human) and a "Means of Motion" (Gear).
>
> 3. **The Risk of "Conducteur":** This is a dangerous False Friend because it is a valid technical word (Electrical), just the *wrong* technical word (Mechanical).