Case Study L3-003
Semantic Integrity in Patent NMT
Polysemy Hallucination & Term Consistency
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	L3-003
Category:	Semantic Integrity — Polysemy
Focus:	Consistency Drift ("Pace Pulse")
Model:	Generic NMT
Domain:	Medical Devices / Electrophysiology

1 The Context: Domain Polysemy

In patent translation, words often have multiple meanings ("polysemy") depending on the technical field. The term "Pulse" is a classic trap:

- **Physiology Context:** A "Pulse" refers to the rhythmic throbbing of arteries (French: *Pulsation*).

- **Electronics Context:** A "Pulse" refers to a transient change in voltage or current (French: *Impulsion*).

In **EP4142594** (Pacemaker Technology), the term "Pace Pulse Signal" refers strictly to the **electrical trigger** generated by the device, not the biological heartbeat of the patient.

> **Key Concept**
>
> **The "Schizophrenic Claim":**
> A critical failure occurs when an AI model oscillates between these two meanings within the **same claim**. This violates the **Rule of Consistency**, suggesting that the claim refers to two different physical phenomena when it only refers to one.

2 The Glitch: The Semantic Split

In Claim 1 of the dataset, the generic model hallucinates two different translations for the exact same antecedent ("pace pulse signal").

2.1 Forensic Evidence (Claim 1)

2.2 Why This Matters

- **Indefiniteness (Art. 84 EPC):** The claim introduces a "signal de pulsation" in the preamble but refers to a "signal d'impulsion" in the steps. Legally, these are distinct terms. The steps technically fail to reference the object defined in the preamble.

Source Segment (English)	NMT Output (French)	Semantic Domain
Preamble: "...detecting a **pace pulse signal**..."	"...détection d'un **signal de pulsation**..."	Biological (Wrong)
Step (e): "...endpoint of the **pace pulse signal**."	"...point d'extrémité du **signal d'impulsion**..."	Electrical (Correct)

Table 1: Inconsistency Hallucination within a Single Claim

- **Technological Distortion:** "Pulsation" implies the device is monitoring blood flow mechanics. "Impulsion" implies it is monitoring electrical voltage. The translation fundamentally alters the description of the invention's sensor array.

3   Alignment Methodology

3.1   Context-Aware Disambiguation

To resolve polysemy errors, we implement a **Semantic Domain Locking** protocol.

> **Alignment Methodology**
>
> **Annotation Process:**
>
> 1. **Domain Tagging:** The entire document is tagged with the primary domain `Physics > Electricity > Pulse_Technique`.
>
> 2. **Exclusion Rules:** We inject negative constraints prohibiting biological equivalents for electrical terms (e.g., `Pulse != Pulsation` when context = `Voltage/Signal`).
>
> 3. **Global Consistency Check:** An automated post-processing script verifies that every instance of "Pace Pulse" maps to the identical French token ("*Impulsion de stimulation*").

3.2   The Corrected Alignment

- **Source:** "pace pulse signal"

- **Target (Invariant):** "*signal d'impulsion de stimulation*"

By enforcing this invariant, we ensure the legal scope of the patent remains focused on the **device's electrical output**, preventing the claim from drifting into unintended physiological monitoring methods.

4   Key Insights

> **Key Concept**
>
> **What This Case Study Demonstrates:**
>
> 1. **AI Lacks "Global Memory":** Generic models translate sentence-by-sentence. They do not "remember" that they translated "Pulse" as "Pulsation" in line 1 when they reach line 10.
>
> 2. **The Danger of Plausibility:** Both "Pulsation" and "Impulsion" are valid French words. A non-technical proofreader might miss the distinction, but a patent examiner will flag the inconsistency immediately.
>
> 3. **Semantic Integrity = Legal Safety:** Alignment is not just about vocabulary; it is about maintaining a coherent definition of the invention throughout the document.