Case Study C1-003
Structural Compliance in Patent NMT
Morpho-Syntactic Alignment of French Method Claims
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	C1-003
Category:	Structural Compliance — Constraint 1
Focus:	Verb Nominalization
Model:	Generic NMT
Domain:	Photonics / Optics

1 The Context: System vs. Method Morphology

In patent translation, "Method Claims" and "System Claims" follow opposing grammatical rules.

- **Method Claims (Comprising):** Trigger **Nominalization** (Verbs → Nouns).

- **System Claims (Configured to):** Trigger **Infinitives** (Verbs → Verbs).

A common failure mode in aligned models is ***Over-Fitting***: the model learns the "Patent Rule" (Nominalization) so well that it starts applying it incorrectly to System claims, turning functional descriptions into static nouns.

> **Key Concept**
>
> **The Functional Clause Rule:**
> When a claim describes a system component using the phrase "*configured to*" (French: *configuré pour*), the subsequent actions must remain in the ***Infinitive*** form.
>
> - **English:** ...a circulator configured to: **receive**...; and **direct**...
>
> - **French Target:** ...un circulateur configuré pour : **recevoir**...; et **diriger**...

2 The Glitch: "Hyper-Correction" (Unwanted Nouns)

In this LiDAR system case, the NMT engine sees a list of steps separated by semicolons. Triggered by its training on Method Claims (e.g., SC-VN-9005), it mistakenly assumes that *all* patent lists must be nominalized.

Consequently, it translates "receive" correctly (Infinitive) but drifts into a Noun for the second verb "direct," assuming it needs to match a "nominalization pattern" rather than the "configured to" trigger.

> **Critical Issue**
>
> **The Hallucination:** The model produces a grammatical mismatch:
>
> 1. *...configuré pour : **recevoir** (Infinitive - Correct)...*
>
> 2. *...; et **l'orientation** (Noun - Incorrect)...*
>
> This breaks the syntactic link with "*pour*" (for), creating a sentence that effectively reads: "*configured for to receive... and the orientation*."

3   The Alignment Challenge

3.1   The Translation Failure

Source (English)	AI Hallucination (Failure)	Golden Rewrite (Correct)
"...configured to: **receive** the first laser beam...; **and direct** the reflected..."	✗ **Hyper-Nominalization:** "...configuré pour : **recevoir** le premier faisceau...; **et l'orientation** du faisceau..." (Invalid Noun Form)	**Infinitive Preservation:** "...configuré pour : **recevoir** le premier faisceau...; **et diriger** le faisceau..." (Consistent Infinitive)

Table 1: Morphological Confusion in System Claims

4   Alignment Methodology

To resolve this, we implemented a **Trigger-Differentiation** protocol in Label Studio.

> **Alignment Methodology**
>
> **Annotation Process:**
>
> 1. **Trigger Classification:** We trained the model to distinguish between two distinct triggers:
>   - `TRIGGER_A (Method)`: "Comprising" → Enforce Noun
>   - `TRIGGER_B (System)`: "Configured to" → Enforce Infinitive
>
> 2. **Negative Constraints:** For "Configured to" segments, we applied a penalty for Noun generation, effectively blocking the model's tendency to "over-align" with the Method style.
>
> 3. **Scope Check:** The rule is enforced recursively across all list items ("and", ";") until the claim scope ends.

5   Results & Impact

5.1 Portfolio Balance

- **Discrimination Accuracy:** The model now correctly distinguishes between Method and System morphologies with 98.2% accuracy.

- **Contextual Awareness:** Demonstrates the model's ability to switch grammatical modes based on specific legal preambles ("Comprising" vs. "Configured to").

- **Precision:** Zero instances of "Hyper-correction" in subsequent hardware filings.