Case Study C1-002
Structural Compliance in Patent NMT
Morpho-Syntactic Alignment of French Method Claims
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	C1-002
Category:	Structural Compliance — Constraint 1
Focus:	Verb Nominalization
Model:	Generic NMT
Domain:	Photonics / Optics

1 The Context: Semantic vs. Structural Accuracy

In Neural Machine Translation (NMT), **Semantic Integrity** ensures the meaning is preserved, but **Structural Compliance** ensures the text adheres to the rigid grammatical frameworks of a specific domain.

While standard language allows for flexible verb forms, French Patent Law requires strict grammatical agreement between the preamble and the procedural steps. This is a critical failure point for generic engines when processing **Method Claims** (*Revendications de procédé*).

> **Key Concept**
>
> **The Legal Requirement:**
> In French patent practice, when a method claim uses the transition word *comprenant* ("comprising"), all subsequent method steps must be **nominalized**—converted from verb forms into substantive (noun) forms.
> This is not a stylistic preference but a **mandatory structural constraint** under European Patent Convention (EPC) Article 84, which requires claims to be "clear and concise."

2 The Glitch: The "Infinitive Drift"

Generic NMT models often translate English gerunds (e.g., "directing") directly as French infinitives (*diriger*).

The transition word *comprenant* strictly dictates that all following list items must be nominalized (transformed into substantives). While the infinitive form is semantically understandable, it creates a **syntax error** in the official filing, rendering the claim grammatically non-compliant.

2.1 Why This Matters

- **Legal Validity:** Patent examiners at the European Patent Office (EPO) will issue objections under EPC Art. 84 for improper verb forms in method claims

- **Claim Clarity:** Infinitive constructions create ambiguity about whether the claim defines a static scope of protection or prescribes procedural instructions

- **Prosecution Delays:** Structural non-compliance can delay patent grant by 12–18 months while corrections are made through the formal amendment process

- **Litigation Risk:** Non-standard grammatical structures provide grounds for claim construction disputes that can weaken enforceability

3   The Alignment Challenge

3.1   The Translation Failure

Source (English)	AI Hallucination (Failure)	Golden Rewrite (Correct)
"A method ... comprising: directing the light..."	✕ **Syntax/Morphology Error:** "...comprenant: diriger la lumière..." (Infinitive Form — Invalid in Method Claims)	**Compliant Translation:** "...comprenant: l'orientation de..." (Substantive/Nominalized Form)

Table 1: Structural Compliance Failure in Method Claim Translation

3.2   The Statistical Bias

Generic NMT models default to infinitive constructions because:

1. Infinitives are statistically more common in general French text

2. The model's training corpus is dominated by conversational and literary French, not legal French

3. Gerund-to-infinitive mapping is the "path of least resistance" in probabilistic decoding

4. The model lacks awareness of domain-specific grammatical constraints

This creates a systematic bias where the model optimizes for **linguistic fluency** rather than **legal formalism**.

4   Alignment Methodology

4.1   Relation Extraction Protocol

To permanently resolve this structural drift, we utilized a **Relation Extraction** workflow in Label Studio.

> **Alignment Methodology**
>
> **Annotation Process:**
>
> 1. **Trigger Identification:** Annotators explicitly identify *comprenant* as the governing trigger word
>
> 2. **Dependency Linking:** Create explicit causal link between trigger and subsequent verb phrase: `comprenant` → `ENFORCES_NOMINALIZATION` → `verb_phrase`
>
> 3. **Error Flagging:** Tag infinitive forms like *diriger* as structural violations
>
> 4. **Correct Mapping:** Provide the nominalized form *l'orientation de* as the gold standard output
>
> This teaches the model that the presence of *comprenant* in a method claim preamble necessitates a grammatical transformation, aligning the engine with French patent filing standards.

4.2 Training Pipeline

1. **Data Collection:** Extract 150+ method claims with nominalization errors from real patent translation projects

2. **Expert Annotation:** Subject Matter Experts (patent translators) manually correct each instance and annotate the trigger-dependency relationship

3. **Pattern Encoding:** Create training pairs:

   - Input: Source claim + NMT output with infinitive
   - Expected: Corrected output with nominalization
   - Annotation: Explicit relation extraction markup

4. **Fine-Tuning:** Apply Reinforcement Learning from Human Feedback (RLHF) with penalty signals for infinitive violations

5. **Validation:** Test on held-out patent claims to measure nominalization accuracy

4.3 Morphological Transformation Rules

The alignment protocol encodes systematic verb-to-noun transformation patterns:

Verb Type	Infinitive	Nominalized Form
-er verbs (general)	*diriger*	*l'orientation de / le guidage de*
-er verbs (-age)	*chauffer*	*le chauffage de*
-ir verbs (-issement)	*refroidir*	*le refroidissement de*
-re verbs	*conduire*	*la conduite de*
Irregular	*transmettre*	*la transmission de*

Table 2: Systematic Nominalization Patterns

5 Results & Impact

5.1 Quantitative Improvement

After fine-tuning on the annotated Gold Set corpus:

- **Nominalization Accuracy:** 96.8% (up from 58.3% baseline)

- **False Positive Rate:** 1.2% (incorrect nominalization of non-method claims)

- **Training Corpus Size:** 187 annotated claim pairs

- **Validation Set Performance:** 94.1% on unseen photonics patents

5.2 Practical Impact

- **Zero EPC Art. 84 objections** for nominalization errors in 43 subsequent filings

- **Prosecution time reduction:** Average 4.7 months faster grant

- **Client confidence:** Elimination of costly post-filing amendments

- **Scalability:** Model generalizes to other technical domains (medical devices, telecommunications)

6 Key Insights

> **Key Concept**
>
> **What This Case Study Demonstrates:**
>
> 1. **Domain-Specific Constraints Override Statistical Preferences:** Generic models optimize for fluency; aligned models enforce legal formalism
>
> 2. **Explicit Relation Extraction Is Essential:** Simply showing correct examples is insufficient—the model must understand *why* a transformation is required
>
> 3. **Structural Compliance Requires Binary Enforcement:** Unlike semantic nuances, grammatical rules are absolute and can be validated programmatically
>
> 4. **HITL Annotation Quality Matters:** Patent-savvy annotators who understand both linguistic structure and legal requirements are critical

7 Related Case Studies

- **SC-VN-9002:** Medical Device Sterilization Method — Nominalization in FDA submission context

- **SC-VN-9003:** Semiconductor Fabrication Process — Complex multi-step nominalization chains

- **SC-AA-9001:** Antecedent Basis Failures — "The at least one" article preservation

- **SI-TL-9001:** Term-Level Hallucination — "Current" polysemy in semiconductor vs. fluidic contexts