

Case Study C4-001

Structural Compliance in Patent NMT

Syntactic Linearity & Modification Scope
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID: C4-001
Category: Structural Compliance — Constraint 4
Focus: Syntactic Linearity Bias
Model: Generic NMT
Domain: Medical Devices / Cardiology

1 The Context: Head-Final vs. Head-Initial Syntax

English technical nomenclature is predominantly **Head-Final**, where modifiers precede the noun (e.g., "*Electrocardiogram Lead*"). French, conversely, is **Head-Initial**, requiring the noun to lead the phrase followed by modifiers (e.g., "*Dérivation d'électrocardiogramme*").

This structural inversion is not merely stylistic; it defines the **modification scope** of the patent claim. In complex noun phrases, failing to invert the structure often breaks the logical link between the component and its function.

Key Concept

The "Linearity Trap":

Generic NMT models process text sequentially (left-to-right). When encountering complex compound terms, they exhibit **Linearity Bias**, translating tokens in their English order without performing the necessary syntactic rotation. This results in a translation that is "word-for-word" accurate but structurally nonsensical.

2 The Glitch: Scope Distortion

In Case Study C4-001 (Med-Tech Patent), the generic model failed to identify "Lead" as the head noun. Instead of restructuring the phrase, it treated the comma-separated modifiers as a list of distinct items.

2.1 Why This Matters

- **Invalidity (Indefiniteness):** The claim describes an "impossible object" composed of three unrelated items (an ECG machine, an acronym, and a lead) rather than a single specific component.
- **Added Subject Matter:** By creating a list where none existed in the source, the translation introduces "New Matter," potentially violating Article 123(2) EPC.

- **Loss of Protection:** The specific invention (the lead itself) is no longer properly claimed, leaving the core technology unprotected.

3 The Alignment Challenge

3.1 The Translation Failure

Source (English)	AI Hallucination (Failure)	Golden Rewrite (Correct)
"...from each of at least one electrocardiogram, ECG, lead... "	<p>× Linearity Error:</p> <p>"...d'au moins un électrocardiogramme, ECG, dérivation..."</p> <p>(List of 3 items — No syntactic inversion)</p>	<p>Compliant Translation:</p> <p>"...d'au moins une dérivation d'électrocardiogramme (ECG)..."</p> <p>(Head-Initial Inversion)</p>

Table 1: Syntactic Linearity Failure in Compound Term Translation

3.2 The Statistical Bias

The model translated the tokens sequentially:

- 1. Electrocardiogram → *Électrocardiogramme*
- 2. ECG → *ECG*
- 3. Lead → *Dérivation*

This "Linearity Bias" occurs because the model attends to local token probabilities rather than the global syntactic tree of the noun phrase.

4 Alignment Methodology

4.1 Dependency Parsing Protocol

To resolve linearity errors, we implement a **Syntactic Reordering** workflow in our post-editing layer.

Alignment Methodology

Annotation Process:

1. **Head Detection:** Identify the true Head Noun (e.g., "Lead") at the end of the English string.
2. **Scope Mapping:** Map all preceding nouns ("Electrocardiogram", "ECG") as adjectival modifiers belonging to the Head.
3. **Inversion Rule:** Enforce the Noun + [De] + Modifier structure in French.
4. **Acronym Handling:** Ensure parenthetical placement of acronyms (ECG) rather than treating them as list items.

5 Key Insights

Key Concept

What This Case Study Demonstrates:

1. **Syntax > Semantics:** Getting the words right is useless if the structure is wrong. "Perfect" vocabulary can still result in an invalid patent.
2. **The "Invisible" Error:** These errors are dangerous because they look correct to non-experts (all the right words are present). [cite_s tart]
2. **Logical Integrity:** Alignment is about preserving the *logic* of the invention (One component vs. Three components), not just the language.

Portfolio: Patent Translation AI Alignment Framework

Author: Cédric Stéphany

Specialization: Technical Translation (FR↔EN) — Patents, Telecommunications, Semiconductors

Contact: cedric@tmcwx.com

Last Updated: January 27, 2026