

Case Study C3-001

Terminology Standardization vs. Grammatical Fluency

The "User Equipment" Trap (Compound Nouns vs. Genitive)
Cédric Stéphany — Technical Translation & AI Alignment Specialist

Case Study Metadata

Dataset ID:	C3-001
Category:	Terminology Enforcement — Level 3
Focus:	Compound Nouns vs. Genitive Construction
Model:	Generic NMT
Domain:	Telecommunications (3GPP/5G)

1 The Context: Fixed Compounds in Telecommunications

In the domain of 3GPP/5G standards, certain terms function as fixed proper nouns rather than descriptive phrases. The term "User Equipment" (UE) defines a specific class of device in the network architecture. While standard French grammar typically requires a preposition to link two nouns (*Nom + de + Nom*), technical nomenclature often relies on **juxtaposition** to designate specific standard entities.

Key Concept

The Standardization Rule:

In 3GPP specifications, "User Equipment" is a monolithic entity. It is not merely "equipment belonging to a user," but a defined network node. Therefore, the French translation must be the fixed compound *équipement utilisateur*, mirroring the English juxtaposition, rather than a grammatical description.

1.1 Why This Matters in Patent Translation

In telecommunications patents referencing 3GPP standards, "User Equipment" must maintain absolute terminological consistency with official specifications. Using non-standard translations like *équipement de l'utilisateur* creates:

- **Standard Misalignment:** Breaks consistency with 3GPP/ETSI terminology databases
- **Acronym Confusion:** The translation must align naturally with "UE" abbreviation
- **Prior Art Searches:** Non-standard terms make it harder to find relevant prior art during examination
- **Claim Interpretation:** Examiners may question whether different terms refer to different components

2 The Glitch: The "Genitive Insertion" Bias

Generic NMT engines interpret "User Equipment" through the lens of general language fluency. The model perceives a possession or qualification relationship and hallucinates a preposition to smooth the syntax for a general reader.

During the Machine Translation Post-Editing (MTPE) process, the generic engine oscillated between two incorrect variants:

1. *équipement de l'utilisateur* (Specific possession - "the user's equipment")
2. *équipement d'utilisateur* (General qualification - "user equipment" as description)

Critical Issue

The Terminological Violation:

While grammatically valid in conversational French, these are **terminological violations** in a patent claim, where "UE" must map strictly to the standardized term to ensure consistency with prior art and technical standards.

The model prioritizes linguistic fluency over domain-specific terminology conventions, treating "User Equipment" as a descriptive phrase rather than a fixed technical term.

3 The Alignment Challenge

3.1 The Translation Failure

Source (English)	AI Hallucination (Failure)	Golden Rewrite (Correct)
"...method used in user equipment , UE..."	<p>× Grammatical Smoothing:</p> <p>"...dans un équipement de l'utilisateur, UE..."</p> <p>OR</p> <p>"...dans un équipement d'utilisateur, UE..."</p> <p>(Genitive Insertion - Non-Standard)</p>	<p>✓ Standard Compliance:</p> <p>"...dans un équipement utilisateur, UE..."</p> <p>(Fixed Compound Noun)</p>

Table 1: Terminology Alignment Failure: User Equipment

3.2 The Statistical Bias

Why does the model insert the genitive construction?

1. **Training Corpus Dominance:** In general French text, noun-preposition-noun constructions (*Nom + de + Nom*) are far more common than direct juxtaposition

2. **Linguistic Naturalness:** The model optimizes for "natural-sounding" French rather than technical nomenclature compliance
3. **Lack of Domain Awareness:** The model doesn't recognize "User Equipment" as a standardized technical term requiring fixed translation
4. **Context Blindness:** Even with "UE" acronym in proximity, the model fails to recognize the terminological constraint

4 Alignment Methodology

To resolve this, we utilized a **Dictionary Match with Negative Constraint** workflow in Label Studio.

Alignment Methodology

Annotation Process:

1. **Entity Tagging:** "User Equipment" is tagged as a DOMAIN_ENTITY (3GPP/Telecommunications)
2. **Negative Constraint:** We explicitly flag the token sequence "*équipement de*" as a FORBIDDEN_PATTERN when triggered by this specific source term
3. **Compound Enforcement:** The model is fine-tuned to recognize that in telecommunications standards, noun juxtaposition (*équipement utilisateur*) overrides standard grammatical genitives
4. **Acronym Validation:** Cross-reference with "UE" acronym to ensure terminological consistency
5. **Standard Database Alignment:** Validate against 3GPP TS 38.300 and ETSI terminology databases

This forces the model to treat "User Equipment" as a single token block rather than a phrase to be grammatically parsed.

4.1 Training Pipeline

1. **Corpus Collection:** Extract 300+ telecommunications patent claims containing "User Equipment" from 5G/LTE patent families
2. **Error Identification:** Flag all instances where generic NMT produced genitive constructions
3. **Expert Correction:** Patent translators with telecommunications expertise provide standardized translations
4. **Dictionary Integration:** Build forced translation lookup:
 - Source: "User Equipment" / "UE"
 - Target (FR): "*équipement utilisateur*"
 - Domain: 3GPP, Telecommunications

- Priority: OVERRIDE grammatical preferences
5. **Fine-Tuning with RLHF:** Reinforcement Learning from Human Feedback with penalty signals for genitive insertion
 6. **Validation:** Test on held-out 5G patents to measure standardization compliance

5 Results & Impact

5.1 Quantitative Improvement

After implementing terminology enforcement:

- **Standardization Compliance:** 99.4% (up from 23.7% baseline)
- **Genitive Insertion Rate:** 0.6% (down from 76.3%)
- **Training Corpus Size:** 342 annotated claim pairs
- **Validation Set Performance:** 98.9% on unseen telecommunications patents

5.2 Practical Implications

- **Standard Compliance:** Ensures full alignment with 3GPP/ETSI terminology databases
- **Acronym Consistency:** The translation *équipement utilisateur* aligns naturally with the abbreviation "UE" (User Equipment), whereas *équipement de l'utilisateur* (EdU) would create a mismatch
- **Zero Post-Editing:** Eliminates repetitive corrections of high-frequency terms throughout the specification
- **Portfolio Consistency:** Same terminology used across client's entire 5G patent family (73 patents)
- **Examiner Confidence:** No terminology-related objections in 89 subsequent filings

5.3 Domain Generalization

The same methodology successfully enforced standardized compounds in other telecommunications terms:

Technical Term	Wrong (Genitive)	Correct (Compound)
"Base Station"	station de base	station de base (accepted)
"Access Network"	réseau d'accès	réseau d'accès (accepted)
"Radio Bearer"	support de radio	support radio
"Control Plane"	plan de contrôle	plan de contrôle (accepted)
"Data Plane"	plan de données	plan de données (accepted)

Table 2: Standardized Telecommunications Compounds

Note: Some terms like "station de base" are standardized *with* the genitive construction in French 3GPP specifications. The key is consistency with the official standard, not blind removal of genitives.

6 Key Insights

Key Concept

What This Case Study Demonstrates:

1. **Domain Standards Override Linguistic Preferences:** Technical nomenclature follows standardization bodies, not general grammar rules
2. **Terminology Consistency Requires Active Enforcement:** Models won't naturally learn fixed compounds from general training data
3. **Acronym Alignment Is Critical:** Translations must consider how they abbreviate and relate to standard acronyms
4. **Context Awareness Is Insufficient:** Even with clear domain markers ("5G," "3GPP," "UE"), statistical preferences dominate without explicit constraints

7 Related Case Studies

- **C3-002:** "Access Point" Compound Enforcement in Wi-Fi Standards
- **C3-003:** ISO 9000 Terminology Compliance in Quality Management Patents
- **L1-001:** "1-Hot" Hallucination — Polysemy resolution in digital logic
- **C1-001:** Verb Nominalization in French Method Claims

Portfolio: Patent Translation AI Alignment Framework

Author: Cédric Stéphany

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

Last Updated: January 6, 2026