

## Case Study L1-015

### Symbolic Hallucination & Token Protection

The "Missing Variable" Trap (Mathematical Variable vs. Indefinite Article)

Cédric Stéphany — Technical Translation & AI Alignment Specialist

#### Case Study Metadata

**Dataset ID:** L1-015  
**Category:** Semantic Precision — Level 1  
**Focus:** Variable Erasure / Linguistic Interference  
**Model:** Generic NMT  
**Domain:** Plastic Optical Fibers (POF)

## 1 The Context: Variable Definitions in Claims

In high-precision manufacturing patents, structural dimensions are often assigned alphabetic variables (e.g.,  $A$ ,  $B$ ,  $d$ ) to be used in mathematical inequalities that define the scope of the invention.

## 2 The Glitch: Symbol-to-Grammar Hallucination

The AI fails to distinguish between the character "A" used as a **symbolic identifier** and the word "a" used as a **grammatical article**. The model's language layer "corrects" the math into grammar.

Source (English)	AI Failure (French)	Impact
"...represented by <b>A</b> ( $\mu\text{m}$ )..."	"...représenté par <b>une</b> ( $\mu\text{m}$ )..."	Variable Erasure
"...represented by <b>a</b> ( $\mu\text{m}$ )..."	"...représentée par <b>une</b> ( $\mu\text{m}$ )..."	Formula Disconnect

Table 1: The transformation of variables into indefinite articles.

### 3 Alignment Methodology

#### Alignment Methodology

##### Symbolic Token Protection:

1. **Mathematical Context Detection:** Identify "represented by" followed by a single character and a unit of measure ( $\mu\text{m}$ ).
2. **Token Locking:** Flag single-character tokens ( $A, a, B, x$ ) in close proximity to operators ( $\leq, \geq, =$ ) as NON-TRANSLATABLE.
3. **Syntactic Pass-Through:** Bypass the grammatical agreement engine for locked tokens to ensure the variable survives the transition into the target language.