

# Nuevos caminos de la gestión terminológica en la era de la traducción automática

Goedele De Sterck

Joaquín García Palacios

UNIVERSIDAD DE SALAMANCA

[desterck@usal.es](mailto:desterck@usal.es)

[gpalacios@usal.es](mailto:gpalacios@usal.es)



Investigación realizada dentro del proyecto de investigación NeoGENes. Junta de Castilla y León. SA059G19

# ÍNDICE

- Mediación “especializada” (lingüística, pero no solo)
- Terminología, mediación y traducción → necesidades nuevas
- Diferentes caminos de la traducción → gestión terminológica diferente

- Preguntas → hipótesis de partida
- Experimento → metodología utilizada
- Resultados
- Comentarios  
→ Conclusiones

# TERMINOLOGÍA, MEDIACIÓN Y TRADUCCIÓN



Comunicación de saberes especializados: nuevas necesidades

- Importancia de la terminología y su gestión en la comunicación y la traducción especializadas: competencias comunicativa y traductora
- De la traducción a la traducción automática y la posesición: ¿una nueva forma de mediar y de gestionar la terminología?
- Cambios importantes que ya se han producido (propia evolución de la terminología y de la teoría de la traducción):
  - Asumir la **variación** en terminología → cuestión de concepto
  - **Uso real** de los términos como guía → cuestión metodológica

# DIFERENTES CAMINOS DE LA TRADUCCIÓN

## Gestión terminológica diferente



- Traducción tradicional
- *CAT* (TAO, traducción asistida por ordenador)
  - Memorias de traducción
  - Gestores terminológicos
  - Corpus
- *MT* (estadística)
- *NMT* (TAN, traducción automática neuronal)
  - Datos
  - Posedición

- Diccionarios → Lexicografía especializada
- Bancos de datos terminológicos  
Bases de datos
- Corpus de textos especializados
- Big data  
¿Gestión terminológica? → posedición

# PREGUNTAS QUE NOS HACEMOS

- ¿Gestiona la TAN correctamente la terminología?
- ¿Vehicula adecuadamente el conocimiento experto?
- ¿Es la terminología un obstáculo para la máquina o, al contrario, no supone mayor problema?
- ¿Dónde falla la TAN? ¿Son estos fallos privativos de la máquina? ¿Qué le queda por aprender? ¿Qué trampas tiende al poseedor?
- ¿Cuáles son los puntos fuertes de la TAN? ¿Suponen un valor añadido?
- ¿Varían los resultados de un dominio del saber a otro? ¿Se registran diferencias de calidad?
- ¿Conserva la máquina el registro (de experto a experto)?
- .... etc, etc.

→ Necesitamos hacer **pruebas empíricas** que nos permitan extraer conclusiones que no estén basadas en nuestra intuición  
→ **Experimento: el experto frente a la máquina**

# EXPERIMENTO



- Par de lenguas: **inglés / español**
- ÁMBITOS DE ESPECIALIDAD (5x2): *ciencias matemáticas* (2), *ciencias geológicas* (2), *ciencias médicas* (2 - audiolología y cardiología), *ciencias sociales* (2 - economía y sociología), *ciencias humanas* (2 - localización y traductología)
- Nivel de especialización: **alto (experto a experto)**
- Textos con máxima densidad terminológica y mínimo tratamiento (Ciapuscio, 2002): **abstract / resumen**
- Condiciones:
  1. Que el *abstract* haya sido producido originariamente y directamente en inglés por un científico hispanohablante (para una revista anglófona)
  2. Que el experto traduzca su propio texto al español
  3. Que para realizar la traducción se base únicamente en su conocimiento como experto
  4. Que el mismo texto sea traducido por un programa de TAN (DeepL)
  5. Analizar los resultados
  6. Comentar esos resultados
  7. Extraer conclusiones al respecto

# PAUTAS PARA LA EVALUACIÓN Y LA POSEDICIÓN DE LA TERMINOLOGÍA EN TAN



→ OBJETIVOS: (1) mejorar la calidad de la TAN y (2) concienciar a los poseedores

## MQM - MODELO DE EVALUACIÓN DE LA CALIDAD (Lommel 2015)

### ➤ 6 DIMENSIONES

#### 4. TERM (TERMINOLOGÍA - *terminology*)

4.1. No sigue base de datos específica (*termbase*).

4.2. No sigue la terminología del campo (*domain-terminology*).

4.3. Uso inconsistente de terminología (*term-inconsistency*).

### ➤ 4 NIVELES DE GRAVEDAD: crítico, importante, menor, indiferente

## OTRO MODELO DE EVALUACIÓN (Mesa-Lao 2013)

### ➤ 4 errores comunes:

1. Terminología (*verbos como sustantivos, sustantivos como adjetivos, uso erróneo del término en el contexto, abreviaciones*)

¿SON  
**SUFICIENTES**  
ESTOS  
MODELOS?

# RESULTADOS (1)



## → PUNTOS DÉBILES EN TAN: un intento de sistematización

### 1. TÉRMINOS QUE NO SE USAN

MED *olivocochlear*

TAN **olivococócico** / **olivocócleo**

EXP olivococlear

olivocócleas son r

olivocócicas ...

olivocóclicas ...

### 2. TÉRMINO ELEGIDO NO ES

GEO *surface samp*

TAN muestra

EXP muestras

superficiales represe

representativas ...

### 3. TÉRMINOS SIN TRA

MED *postpa*

TAN **postpac**

EXP post-esti

de superficie ...

de superficies ...

de la superficie ...

significativas ...

### 4. ANÁLISIS SINTAGMA

MED *nicotin*

TAN receptor

EXP receptor

materiales ...

cutáneas ...

regionales ...

en superficie ...

### 5. INCONSISTENCIA

#### a. CON RESPECTO AL CAMPO

MAT *solvability condition*

TAN condición de **solvencia** (FIN)

EXP condición de resolubilidad (MAT)

#### b. A LO LARGO DEL TEXTO

SOC *(social) standards / standard-based methods*

TAN **normas** (sociales) / **métodos basados en estándares / métodos estándar**

EXP estándares (sociales) / métodos basados en estándares

### 6. FRASEOLOGÍA (artículos, preposiciones, colocaciones)

MAT *reduce by four the order of Euler-Lagrange equations*

TAN **reducir a cuatro** el orden de (...)

EXP reducir en cuatro unidades el orden de (...)



# CASO CONCRETO: las siglas



Los puntos anteriores confluyen en el uso de las siglas y los términos desarrollados correspondientes:

- siglas inexistentes,
- siglas poco usadas,
- siglas no traducidas,
- siglas mal elegidas (inconsistencia campo – texto),
- siglas mal combinadas (artículo – término desarrollado).

Estrategia de redacción y traducción recomendada:

- primera aparición → término desarrollado y sigla entre paréntesis
- a continuación → solo sigla

MED	EXP	TAN
The auditory cortex (AC) regulates (...).	La corteza auditiva (CA) regula (...).	La corteza auditiva (CA) regula (...).
(...) AC has a reinforcement role	(...) la CA tiene un papel de refuerzo (...).	(...) la AC tiene un papel de refuerzo (...).
(...) descending inputs from AC are needed (...)	(...) las entradas descendientes desde la CA son necesarias (...)	(...) las entradas descendientes de CA son necesarias (...).
(...) descending connections from AC (...)	las conexiones descendientes desde la CA (...).	(...) las conexiones descendientes de la CA (...).
(...) after unilateral ablation of the AC (...)	(...) después de ablación unilateral de la CA (...).	(...) después de la ablación unilateral del AC (...).
Unilateral AC ablation induced (...)	La ablación unilateral de la CA indujo (...).	La ablación AC unilateral indujo (...).
Unilateral removal of the AC also resulted in (...)	La extirpación unilateral de la CA también produjo (...).	La remoción unilateral del AC también resultó en (...).
(...) AC ablations induced (...)	(...) la ablación de la CA indujo (...).	(...) las ablaciones de CA indujeron (...).

→ Los problemas terminológicos varían de un campo de conocimiento a otro (número y gravedad)

¿Están unos campos más y mejor representados que otros (*input* para la máquina)?

→ No todos los fallos son exclusivos de la máquina: las traducciones realizadas por expertos también presentan problemas terminológicos (incoherencias, lagunas y descuidos).

¿Prisas, falta de hábito, falta de preocupación terminológica y lingüística en español (que no en inglés)...?

---

### ¿Cuál es la versión del experto?

El móvil comienza por completo con nueva velocidad hasta que un nuevo reset ocurre.

El móvil se reinicia desde cero con una nueva velocidad hasta que se produce el siguiente reinicio.

---

→ La terminología más cercana a la lengua común (HUM y SOC) es la que menos problemas presenta.

Analiza la **trayectoria laboral** de las **mujeres migrantes** en el **servicio doméstico**.



La **remoción unilateral del AC** también resultó en un **aumento de la regulación** de las transcripciones para la subunidad **nAChR de  $\alpha 7$** .

# PUNTOS FUERTES en TAN: un intento de sistematización



## RESULTADOS (2)

### → RIGOR CON RESPECTO AL ORIGINAL

#### 1. SIN OMISIONES

MED *episodes terminated or not at ATP-2*

TAN episodios terminados o no en ATP-2

EXP episodios terminados ∅ tras ATP-2

#### 2. RESPETA LAS ESTRATEGIAS DE REDACCIÓN

GEO *the abatement of As (...) enriched in As*

TAN la reducción de la presencia de As (...) se enriquecieron en As

EXP la disminución del arsénico (...) enriquecidos en As

#### 3. RESPETA EL REGISTRO ESPECIALIZADO

HUM *using Swordfish II*

TAN utilizando Swordfish II

EXP que se llevó a cabo en la herramienta de TAO Swordfish II

## COMENTARIOS (2)

En algunos aspectos, la TAN parece ser incluso más sistemática que el experto humano

→ Sobre todo en la dimensión horizontal: de original a traducción

→ Falla la consistencia vertical: campo – texto

# OTROS COMENTARIOS (sobre el experto)



## UN EXTREMO

LOS MÁS CONCIENCIADOS Y LOS MÁS COMPROMETIDOS CON EL ESPAÑOL CIENTÍFICO tienden a aplicar **estrategias divulgativas**

→ **cambio de registro** - de experto a (semi)lego:

- menor densidad terminológica
- mayor tratamiento (uso de comillas, explicaciones, “por las siglas en inglés”, etc)

## OTRO EXTREMO

LOS MENOS CONCIENCIADOS Y LOS MÁS DESPREOCUPADOS POR LA CIENCIA EN ESPAÑOL tienden a **calcar** las convenciones del **inglés científico**

→ **“espanglish” científico**

- anglicismos,
- omisión de artículos y preposiciones,
- etc.

## RESULTADO

(In)conscientemente, unos y otros contribuyen al **debilitamiento del español científico**

# CONCLUSIONES



## Algunas cuestiones fundamentales

1. Importancia de la terminología para vehicular el conocimiento especializado → textos especializados
2. Necesidad de perfeccionar las prestaciones terminológicas de la TAN.
3. Valor fundamental de los textos reales para enseñar a la máquina (**terminología contextualizada y de calidad**)
4. Nunca términos sin contextos (**buenos contextos de uso - dimensión textual**)
5. Corpus cada vez son más importantes (cantidad, **calidad y variedad**)
6. Hacen falta textos en las lenguas románicas (**textos de nivel de especialización alto, ahora solo en inglés**)
7. No solo corpus paralelos, sino también comparables.
8. Concienciar a los expertos: **más allá de la divulgación y del espanglish científico.**
9. Reafirmar los principios básicos de la terminología actual (teoría y actuación)
10. Nuevos caminos, **no tan nuevos...**

→ **RETOS DE FUTURO**

# REFERENCIAS BIBLIOGRÁFICAS



- CIAPUSCIO, G. e I. KUGUEL (2002), “Hacia una tipología del discurso especializado: aspectos teóricos y aplicados”. En J. García Palacios y M. Teresa Fuentes (eds.), *Entre la terminología, el texto y la traducción*, Salamanca, Almar, pp. 37-73.
- GÖRÖG, A. y P. SÁNCHEZ-GIJÓN (2014), *Traducció i qualitat*, núm. 12 de la revista *Tradumàtica. Tecnologies de la traducció*.
- ISABELLE, P. *et al.* (2017), “A Challenge Set Approach to Evaluating Machine Translation”, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, pp. 2486–2496.
- KOPONEN, M. (2016), “Is machine translation post-editing worth the effort? A survey of research into post-editing and effort”, *The Journal of Specialised Translation*, 25, pp. 131-148.
- LOMMEL, A. *et al.* (2015), *Multidimensional Quality Metrics (MQM) Issue Types*. Disponible en: <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>
- MESA-LAO, B. (2013), *Introduction to post-editing – The CasMaCat GUI*. SEECAT project, Copenhagen, Center for Research and Innovation in Translation and Translation Technology.
- TAUS (2016), *TAUS Quality Dashboard. From Quality Evaluation to Business Intelligence*, Países Bajos.
- TORAL, A. *et al.* (2018) “Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation”, *Proceedings of the Third Conference on Machine Translation: Research Papers*, Bruselas, Association for Computational Linguistics, pp. 113-123.

Muito obrigados  
Muchas gracias

Proyecto NeoGENes: el préstamo léxico como mecanismo actualizador del español general y especializado  
Junta de Castilla y León. SA059G19

[gpalacios@usal.es](mailto:gpalacios@usal.es)

[desterck@usal.es](mailto:desterck@usal.es)

