

Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

Moderadora
Ona de Gibert



Text Mining Unit
(BSC)

Ponente
Marta Villegas



Text Mining Unit
(BSC)

Ponente
Raquel Xalabarder



Catedrática Propiedad
Intelectual (UOC)

Ponente
Javier Hernández



Subdirector General de
Desarrollo e
Implantación de
Servicios Digitales
(Ministerio de Justicia)

Ponente
Alicia Ollé



Directora de
operaciones de
cliente (Parlem
Telecom)



Importancia de los datos lingüísticos de calidad para la IA y las TL

ELRC WHITE PAPER

Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe

WHY LANGUAGE DATA MATTERS

Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe - Why Language Data Matters: ELRC White Paper.

November 2019

#elrc_spain2021



in English!
event code: ELRC_SPAIN2021

Importancia de los datos lingüísticos de calidad para la IA y las TL

“Textual data is not perceived as a valuable asset”

ELRC WHITE PAPER

**Sustainable Language Data Sharing
to Support Language Equality
in Multilingual Europe**

WHY LANGUAGE DATA MATTERS



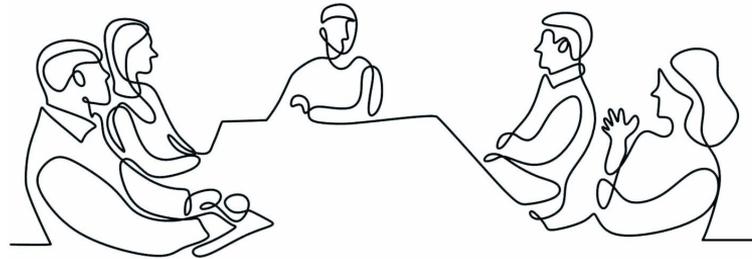
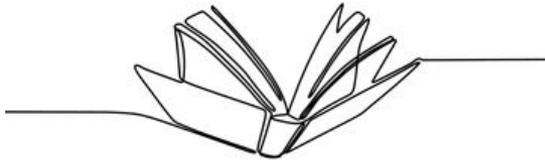
Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan grandes cantidades de datos lingüísticos



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan grandes cantidades de **datos lingüísticos**



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan **grandes cantidades** de datos lingüísticos

Modelo del Lenguaje	Palabras
XLM-R	295.008,00 M
GPT-3	96.755,00 M



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan **grandes cantidades** de datos lingüísticos

Modelo del Lenguaje	Palabras
XLM-R	295.008,00 M
GPT-3	96.755,00 M
Don Quijote	0,38 M



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan grandes cantidades de datos lingüísticos
- Datos lingüísticos de **calidad**
 - Calidad **formal**
 - Calidad de **contenido**



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan grandes cantidades de datos lingüísticos
- Datos lingüísticos de **calidad**
 - Calidad **formal**
 - Calidad de **contenido**

The doctor went out of the
operating room.



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan grandes cantidades de datos lingüísticos
- Datos lingüísticos de **calidad**
 - Calidad **formal**
 - Calidad de **contenido**

The doctor went out of the
operating room.

El doctor salió del quirófano.

Traducción de Google Translate
consultada el 13/05/2021



Importancia de los datos lingüísticos de calidad para la IA y las TL

- Los sistemas de TL necesitan grandes cantidades de datos lingüísticos
- Datos lingüísticos de **calidad**
 - Calidad **formal**
 - Calidad de **contenido**
 - **Privacidad**



Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

Moderadora
Ona de Gibert



Text Mining Unit
(BSC)

Ponente
Marta Villegas



Text Mining Unit
(BSC)

Ponente
Raquel Xalabarder



Catedrática Propiedad
Intelectual (UOC)

Ponente
Javier Hernández



Subdirector General de
Desarrollo e
Implantación de
Servicios Digitales
(Ministerio de Justicia)

Ponente
Alicia Ollé



Directora de
operaciones de
cliente (Parlem
Telecom)



Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

- ¿Existe sensibilidad a nivel político por la necesidad del intercambio y la reutilización de datos lingüísticos?
- ¿Incluye el portal nacional de datos abiertos datos lingüísticos en formato procesable?
- ¿Se menciona explícitamente en la estrategia española de IA?

Ponente
Marta Villegas



Text Mining Unit
(BSC)



Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

- ¿Cómo ha avanzado la legislación española en materia de datos abiertos?
- ¿Se habla explícitamente de datos lingüísticos?
- ¿Hay diferencias entre los datos del sector público y los del sector privado?

Ponente
Raquel Xalabarder



Catedrática Propiedad
Intelectual (UOC)



Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

- ¿En su opinión se fomenta el intercambio de datos lingüísticos en el sector público?
- ¿Cuál es la problemática específica de su Ministerio?
- ¿Existen procedimientos que faciliten poder compartir datos?

Ponente
Javier Hernández



Subdirector General de
Desarrollo e
Implantación de
Servicios Digitales
(Ministerio de Justicia)



Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

- ¿Qué lugar ocupa una empresa como Parlem Telecom en la cadena de valor de los datos?
- ¿Existen procedimientos que faciliten poder compartir datos?

Ponente
Alicia Ollé



Directora de
operaciones de
cliente (Parlem
Telecom)



Panel de discusión: Gestión, intercambio y reutilización de datos lingüísticos en España

¿Qué habría que hacer para mejorar la situación
y facilitar el intercambio y la reutilización de datos lingüísticos?



Ponente
Marta Villegas



Ponente
Raquel Xalabarder



Ponente
Javier Hernández



Ponente
Alicia Ollé

