

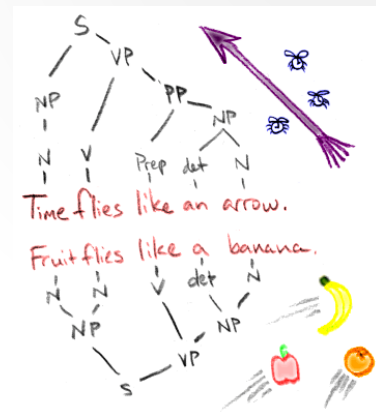
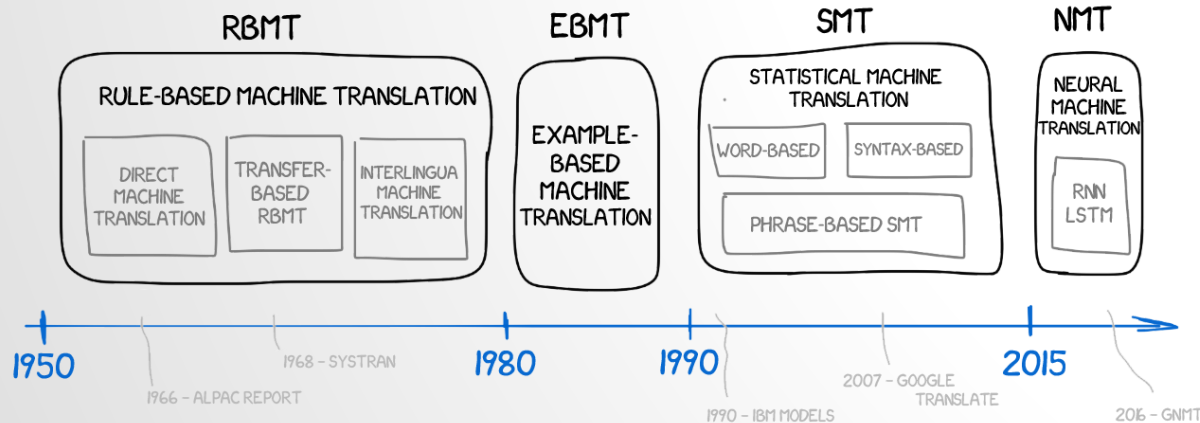
ELRI en el ecosistema de traducción europeo y español

Madrid, 12 de marzo 2019



Hª de la Traducción Automática: De los lingüistas a los datos

A BRIEF HISTORY OF MACHINE TRANSLATION



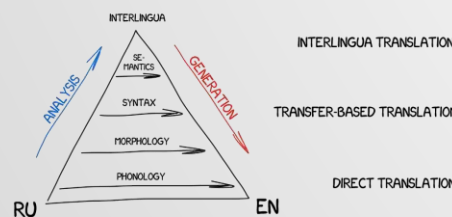
DualiaTeletraducción
@Dualia

OJO a la redacción. Una coma mal puesta convierte a Pablo Alborán en la Drag Queen del verano.
eldiariomontanes.es/culturas/musica ...

Translate Tweet



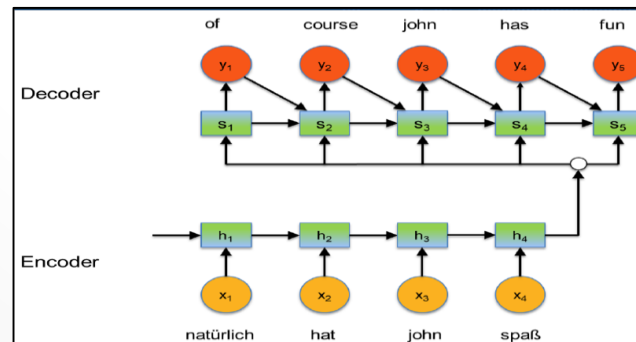
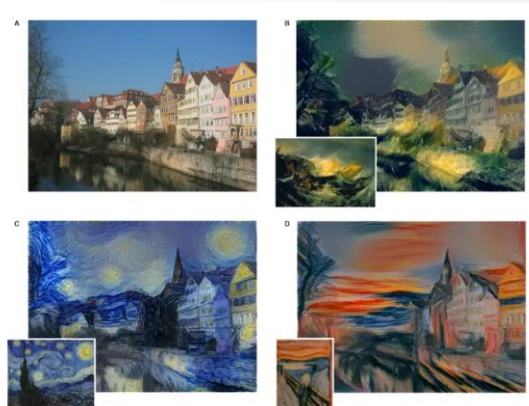
VAUQUOIS TRIANGLE



Fuente: "A history of machine translation from the Cold War to deep learning" [I. Pestov]

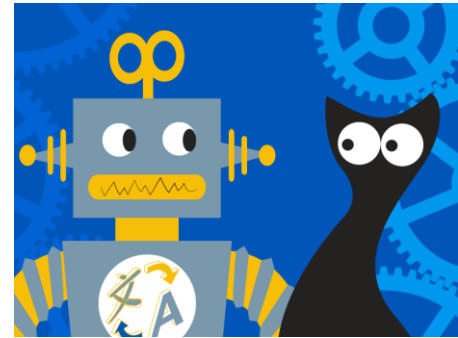
La revolución de las redes neuronales y el aprendizaje profundo en la TA

- La TA neuronal es más **fluida**.
- **Selecciona** mejor entre posibles traducciones de una palabra.
- Puede generar traducciones que **no ha visto nunca** en fase de entrenamiento.
- Traducción de más calidad incluso para pares de lenguas con pocos recursos gracias al **aprendizaje por transferencia**.



La integración de la TA neuronal en el flujo del traductor

- Hay que entender que la TA neuronal y las traducciones humanas son dos cosas diferentes, pero **complementarias**.
- La TAn puede ayudar a hacer traducciones humanas más **eficientes**, sugiriendo alternativas.
- La traducción humana, a su vez, es **imprescindible para proporcionar datos de entrenamiento de gran calidad**.
- Los traductores tenderán a aplicar cada vez más su conocimiento lingüístico en tareas de **postedición**.



Fuente: Ciklopea

Multilingüismo y Mercado Único Digital

- La UE está haciendo grandes esfuerzos para adaptar el Mercado Único europeo a la era digital, a través del desarrollo de **infraestructuras interoperables e interconectadas** que ponen en comunicación a los ciudadanos europeos por encima de las fronteras nacionales y las barreras lingüísticas. Y a la vez:
- Preservar la diversidad lingüística
- Modernizar la Administración Pública
- Aumentar la competitividad de Europa y su independencia de proveedores tecnológicos externos.



Programa de financiación CEF Digital

- Construcción de las **infraestructuras de servicios digitales** (DSI) en varios sectores:
 - Online Dispute Resolution
 - Europeana
 - Business Registers Interconnection System
 - Electronic Exchange of Social Security Information
 - eHealth
 - eJustice
 - European Data portal
- Implementación de servicios transversales (p.e. eID, eDelivery, **eTranslation**...).

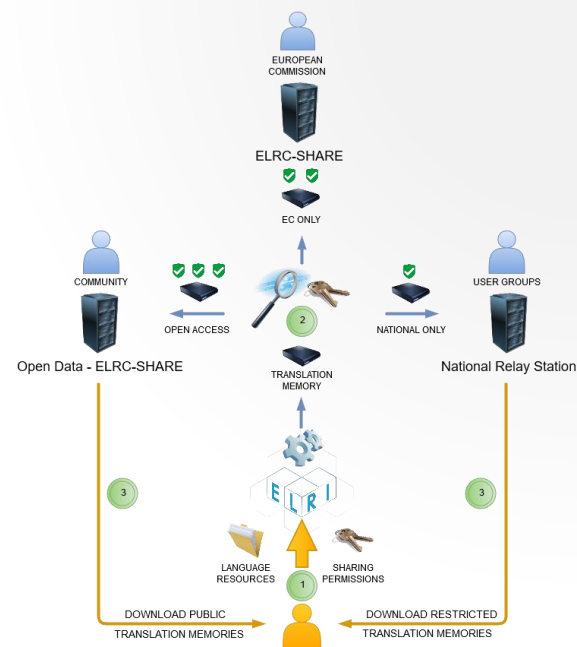


eTranslation + ELRI



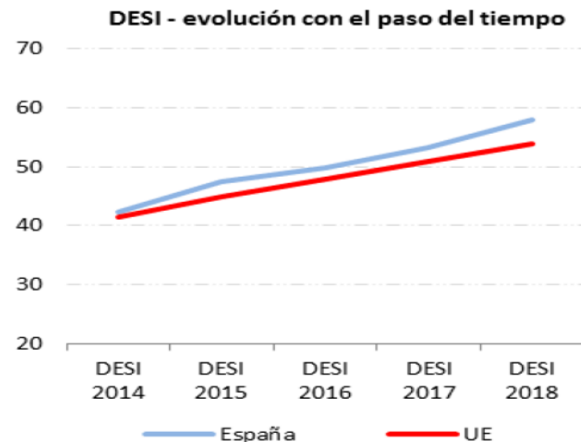
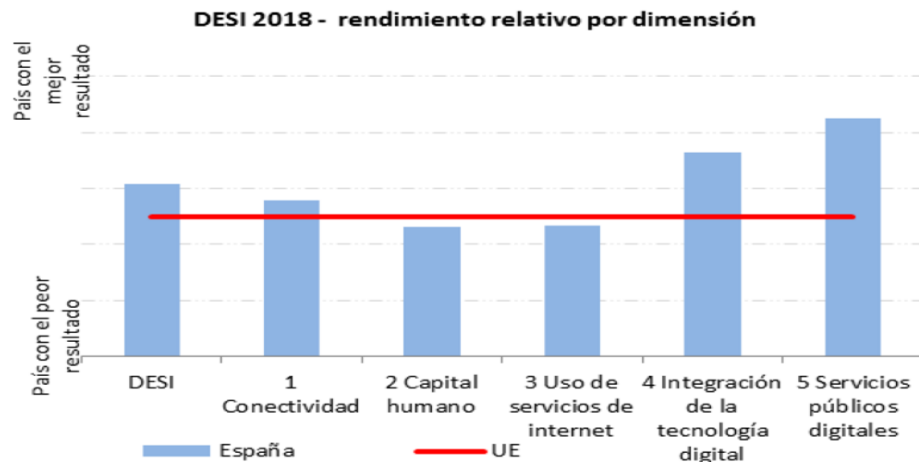
24 lenguas oficiales ⇔ inglés

- Abierto a todas las administraciones públicas.
- Entrenado sobre texto legal y administrativo.
- Para dar servicio a las DSIs se necesitan datos específicos del sector en todas las lenguas de la Unión.



El papel de ELRI en este ecosistema es el de recogida de datos que permitan generar una TA de calidad

España en el Mercado Único Digital

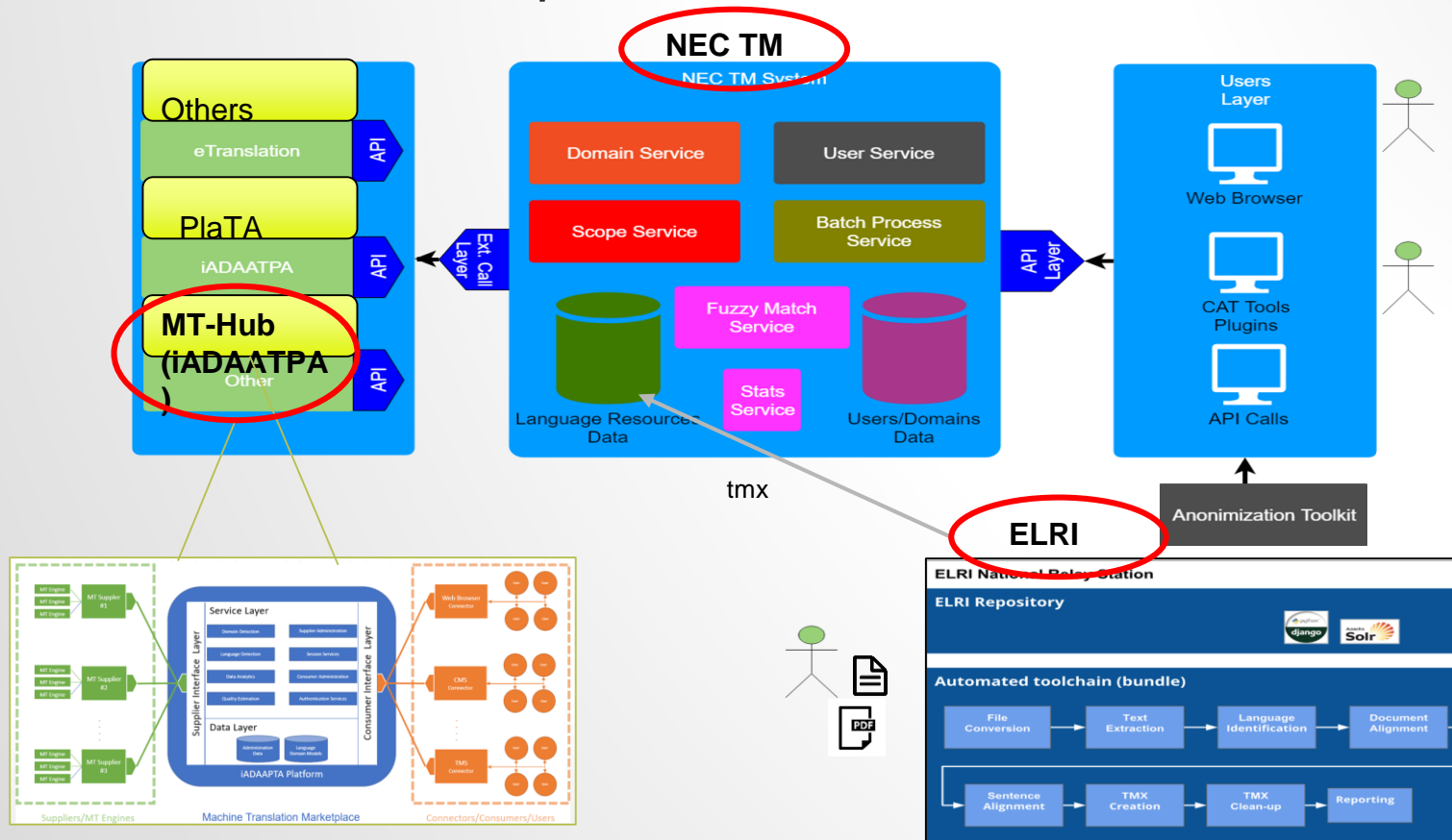


5 Servicios públicos digitales	España		Grupo puntuación	UE puntuación
	puesto	puntuación		
DESI 2018	4	72,4	58,5	57,5
DESI 2017	4	68,5	54,9	53,7

Plan de Impulso de las Tecnologías del Lenguaje



Infraestructura de soporte a las tecnologías de la traducción en España



- Los avances en IA en los últimos 3 años han transformado completamente las Tecnologías de la traducción.
- La TA es una tecnología habilitadora clave para las infraestructuras de servicios digitales de la UE.

- Se necesitan **datos** específicos de dominio para entrenar motores de TA de calidad.
- El papel de ELRI es el de recoger y preparar esos datos.

Gracias