

# Estrategia Formativa en TL

Aitziber Atutxa, Núria Bel, Arantza Díaz de Ilarraza,  
Koldo Gojenola, Raquel Martínez, Patricio Martínez,  
Antoni Oliver, German Rigau, Alfonso Ureña

# Estrategia Formativa en Tecnologías del Lenguaje

- Realizar un **estudio preliminar** para la definición de un Libro Blanco para la formación en tecnologías del lenguaje en el que se diseñe una estrategia formativa en TL específica para el sistema universitario español.

# Plan de Impulso a las TL y formación

- El [Plan TL](#) (octubre 2015) se estructura en cuatro ejes.
- Eje II: Impulso a la Industria de las tecnologías del lenguaje
- Medida II.1.2:
- *“Eleva la formación: Inclusión de asignaturas específicas sobre tecnologías del lenguaje en el CV académico, impulso de la creación de contenidos formativos en línea (MooCs), hackatones, apoyo a programas de doctorado industrial, másteres y becas especializadas”*

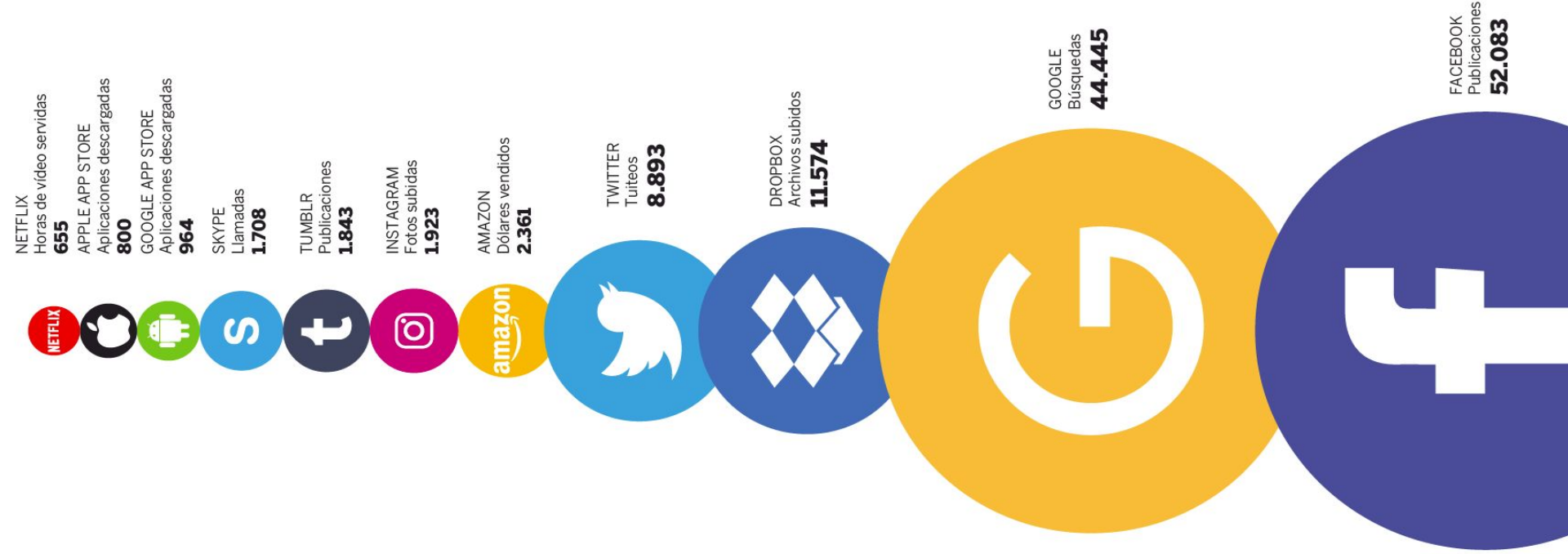
# Plan de Impulso a las TL y formación

- Estudio de Caracterización del Sector de TL en España (octubre 2018):
  - Sector vinculado a la educación superior universitaria.
  - **70%** de los empleados del sector son titulados superiores, doctores o ingenieros
  - **82%** afirma haber contratado personal especializado en TL
  - **28%** han tenido **dificultad** para encontrar los técnicos adecuados
  - Insuficiente oferta de titulados especialistas con formación mixta
  - El informe recomienda:
    - “la creación de una titulación técnica que integre los conocimientos técnicos y de programación, y los conocimientos lingüísticos necesarios para desarrollar la actividad de tecnologías del lenguaje”.

# Big Data y TL

- El contenido digital **no estructurado** representa el **90%** de toda la información [[Libro blanco IDC 2014](#)] ...
- En forma de **textos** (e imágenes, audio y video) en varios **idiomas** ...
- Solo las técnicas apropiadas pueden procesar estos datos ...
- TL están entre las 10 principales tendencias de tecnología analítica estratégica para 2019 según [Gartner](#)

# Qué ocurre en internet cada segundo? (julio 2015) ...



# Qué ocurre en internet cada segundo? (julio 2015) ...



... y no sólo en internet.

- LexisNexis recibe cada día 1,5M noticias ...
- CENDOJ almacena 7M sentencias judiciales (0.6M/año)
- ?M informes médicos (Electronic Health Records, EHR) ...
- ?M de documentos en cada organización ...
- ... y sólo con las **técnicas adecuadas** puede procesarse toda esta información digital no estructurada.



# DEEP LEARNING - A NEW COMPUTING MODEL

## “Training”



ImageNet

LEARNING  
ALGORITHM  
“millions of trillions  
of FLOPS”

TB or PB of  
training data



Known  
Ground Truth  
Labels

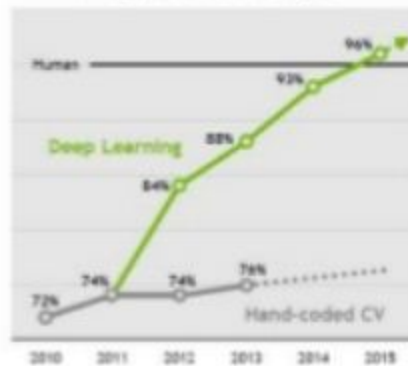
## “Inference”



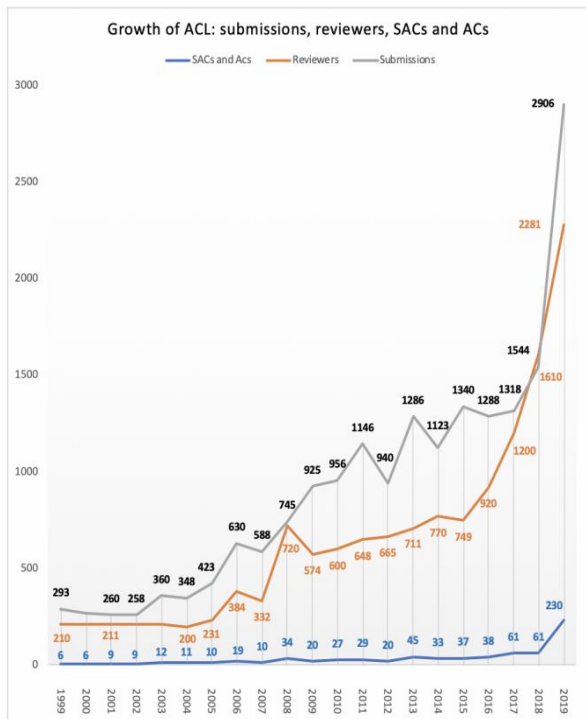
“cat”

## “SUPERHUMAN” RESULTS SPARK HYPERSCALE ADOPTION

ImageNet – Accuracy %



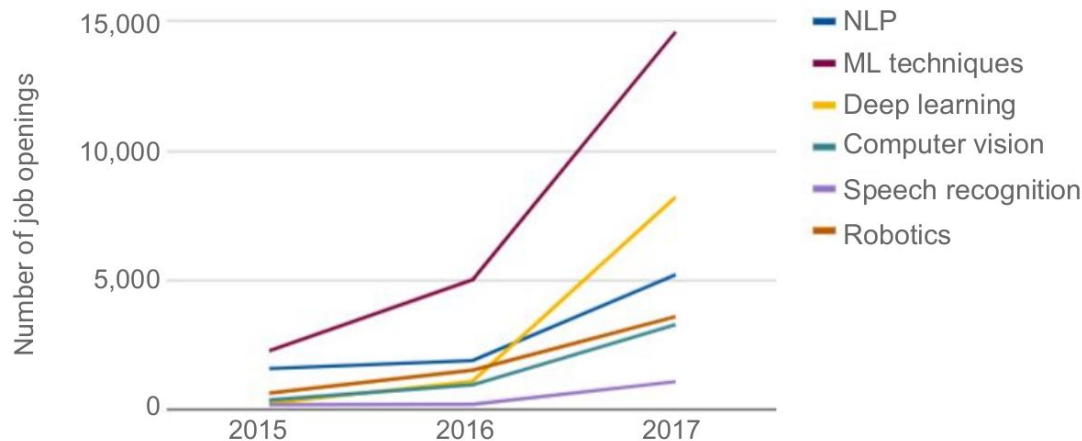
# Interés creciente en TL



Fuente: <http://acl2019pcb.blog.fileli.unipi.it/>

## Job openings by AI skills required (2015 – 2017)

Source: Monster.com

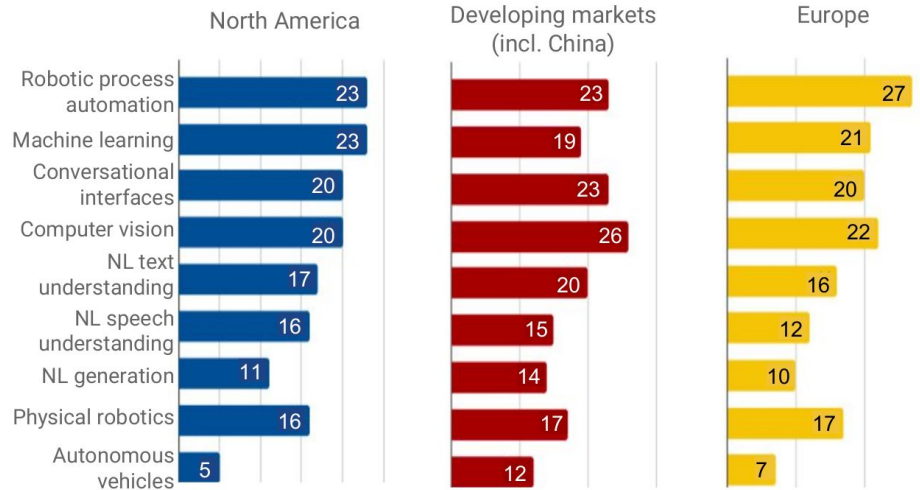


Fuente: [2018 AI Index report](#)

# IA y TL en la industria

Capabilities embedded in at least one company function (2018)

Source: McKinsey & Company



Percent of respondents



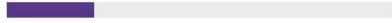





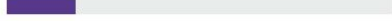
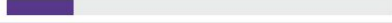
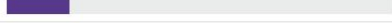
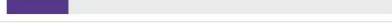
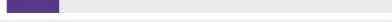
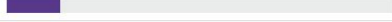
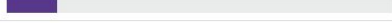
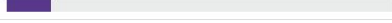
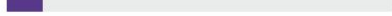
Note: The size of each bar is relative to the capabilities within each region; North America: N = 479; Developing markets (incl. China): N = 189 (China N = 35); Europe: N = 803

Fuente: [2018 AI Index report](#)

# Sesgo de género

## Women represented 18% of the conference authors overall

Spain, Taiwan and Singapore had the highest percentage of female conference authors, but the U.S. accounted for almost half in absolute numbers.

Country	Talent by Gender	Total	Ratio 2019
Spain		170	26%
Taiwan		150	23%
Singapore		310	23%
Australia		550	22%
China		2295	22%
United States		9835	20%
Switzerland		260	19%
Italy		475	18%
United Kingdom		1420	18%
India		580	17%
France		795	16%
Germany		1025	16%
Israel		330	14%
Canada		970	14%
Netherlands		230	13%
South Korea		345	12%
Japan		755	9%

Source: 2019 Global Talent Report • [jf@gagne.ai/talent](https://gagne.ai/talent)

Fuente: [Global AI Talent Report 2019](#)

# Estrategia Formativa en TL

- ET1: Estudio sobre la **oferta formativa** en TL a nivel nacional, europeo e internacional en países de referencia.
- ET2: Propuesta de **acciones formativas** para la estrategia de formación en TL
- ET3: Directrices generales para abordar la estrategia de formación. Identificación de **sectores** para la estrategia de formación TL.
- ET4: Diseño de la **estrategia** de formación en TL específica para el sistema universitario
- ET5: **Plan de acción** a corto y medio plazo para la formación en TL en el sector universitario.

# Análisis previo

- **Informes** existentes
- **Oferta formativa actual** en el ámbito universitario en el área de las TL a nivel nacional, europeo e internacional en países de referencia.
- Estudio de otras **experiencias similares** que hayan abordado la inclusión de nuevas temáticas en los diversos planes formativos (Ciberseguridad).
- **Ofertas de empleo** en el ámbito de las TL
- **Demanda de los empleadores** mediante encuestas.

# Informes existentes en IA y TL

- El [Plan Coordinado en Inteligencia Artificial](#) de la Comisión Europea indica que todos los Estados miembros se enfrentan a la **escasez de profesionales** en el área.
- La [Estrategia Española de I+D+i en IA](#) insta a incluir la IA en el sistema educativo como palanca de cambio tecnológico del país.
- [For a Meaningful Artificial Intelligence Towards a French and European Strategy](#) del parlamento francés (2018), prevé la necesidad de multiplicar por tres en número de expertos en IA en los próximos 3 años. Entre los expertos en IA se menciona especialmente a los **expertos en TL**.

# Informes existentes en IA y TL

- La [Strategic Research, Innovation and Deployment Agenda for an AI PPP](#) desarrolla el tema en "Skills and Knowledge" (página 29) y el Working Area 2 "Skills and Acceptance".
- El informe de la OCDE [Artificial Intelligence in Society](#) desarrolla el tema de en "Preparing for job transformation and building skills".
- El Grupo de expertos de alto nivel de la Comisión Europea sobre IA (AI HLEG) publicó su segundo informe principal, sobre [Recomendaciones de política e inversión para IA confiable](#). El apartado F del capítulo 2 (página 31) está dedicado a "Generando habilidades y educación apropiadas para la IA".



# Oferta formativa actual

- En titulaciones relacionadas con la **lingüística** (filologías, traducción e interpretación, etc.) como en las relacionadas con la **informática** (o telecomunicaciones, ciencia de datos, etc.).
  - Apenas en asignaturas de grado y algunos másteres.
  - No existe ningún grado en TL.
- La formación en TL es **escasa** y **heterogénea**.
- Por ejemplo, tras terminar el grado la inmensa mayoría de estudiantes de informática desconocen cómo procesar adecuadamente la mayor parte de la información digital.

# Demanda de empleo

## Dificultades para encontrar los expertos en TL adecuados

25 respuestas

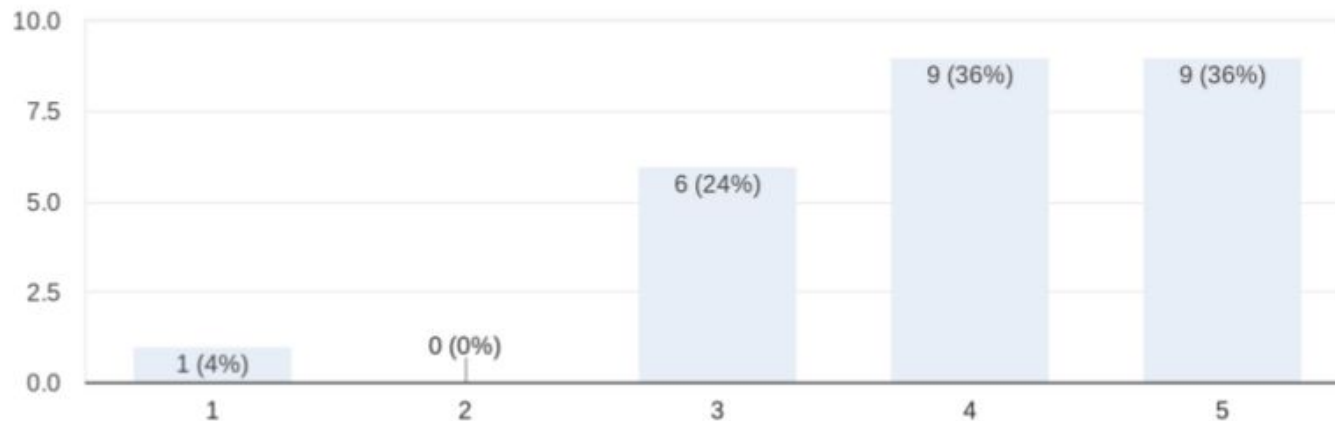


Figura 9. Dificultad para contratar expertos en TL (1: ninguna dificultad, 5: dificultad alta).

# Competencias requeridas en TL

*Tabla 12. Competencias requeridas en empleos de TL.*

Computer Science	2
Artificial Intelligence	1
Big Data	4
Natural Language Processing (NLP)	6
Deep Learning, Neural Nets	11
Machine Translation	6
Machine Learning	6
Speech Processing	5
Chatbots	3
Others: English, blockchain, problem solving, mathematics, graphs, ethical hacking, biomedical processing	

# Demanda de empleo

Contratos de formación de investigadores en empresas (doctorados industriales) y sus ayudas

25 respuestas

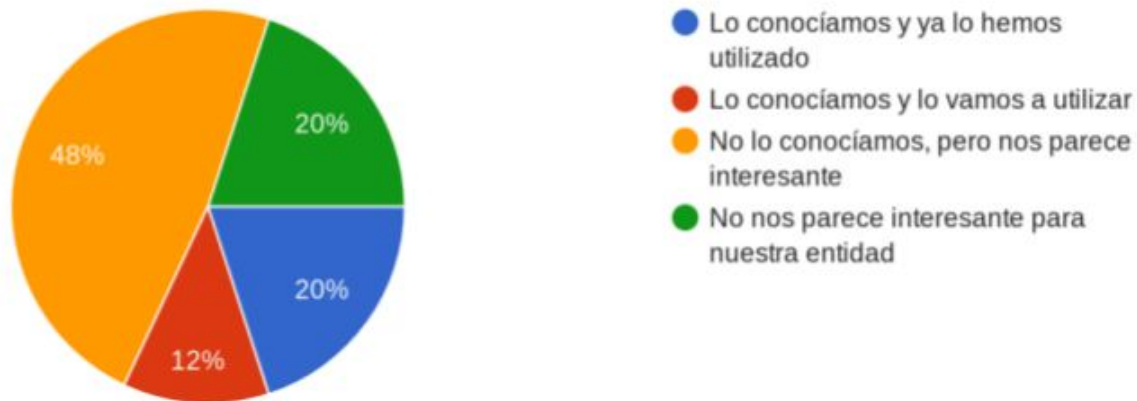


Figura 10. Ayudas a contratos de investigadores en empresas.

# Sectores y grupos objetivo

- Sectores
  - Empresarial
  - Administración pública
  - Ámbito universitario
  - Ciudadanía en general
- Grupos Objetivo
  - Profesionales **expertos** en TL de empresas y administración
  - Profesionales **usuarios** de TL de empresas y administración
  - **Cuadros medios** y profesionales en general de empresas y administración
  - Ámbito universitario
  - Educadores

# Acciones formativas

- Formación continua sobre investigación puntera
- Seminarios y talleres sobre herramientas avanzadas en TL
- Cursos MOOC
- Hackatones
- Aulas universidad-empresa para casos prácticos
- Seminarios y talleres presentando casos de éxito
- Charlas divulgativas
- Formación universitaria: grado, postgrado
- Redes europeas de programas de doctorado
- Jornadas de difusión

# Contenidos específicos

- **Lingüística** para las TL:
  - Comunicación, fonología, morfología, sintaxis, semántica, pragmática, etc.
- **Técnicas** para las TL:
  - basadas en conocimiento, en corpus, estadísticas, aprendizaje automático, deep learning, etc.
- **Aplicaciones** funcionales de las TL:
  - EI, RI, TA, Reconocimiento y Síntesis del Habla, Sistemas de Diálogo, chatbots, etc.
- **Dominios** de aplicación de las TL:
  - medicina, legal, sociología, periodismo, humanidades, ADEs, etc.

# Estrategia formativa universitaria en TL

- Nuevos títulos:
  - Grado en **Lingüística Computacional**, a imagen de los que ya existen a nivel europeo.
  - Grado en **Inteligencia Artificial**, a imagen de los que ya existen a nivel europeo, con líneas curriculares en **Tecnologías del Lenguaje**.



# Estrategia formativa universitaria en TL

- Inclusión de asignaturas obligatorias en **Tecnologías del Lenguaje** y en general de procesamiento de datos no estructurados en el currículum académico de los grados actuales en **Ingeniería en Informática** (y títulos relacionados Ciencia de Datos, etc.) e **Ingeniería de Telecomunicaciones**.
- Inclusión de asignaturas obligatorias en **Tecnologías del Lenguaje** en los grados actuales en **Lingüística** (y títulos relacionados: Lingüística Aplicada, Filología y Lingüística, Traducción e Interpretación, etc.)

# Estrategia formativa universitaria en TL

- Inclusión de contenidos en **Tecnologías del Lenguaje** en grados relacionados con los **dominios de aplicación** de la TL como medicina, sociología, periodismo, derecho, administración y dirección de empresas, etc.

# Estrategia formativa universitaria en TL

- Títulos de **máster específicos** de TL.
- **Programas de doctorado** (incluidos programas de doctorado industrial) y títulos de postgrado (no-oficiales).
- Cursos de **expertos** (< 60 ECTS) para formación continua
- **MOOCs**
- Apoyo efectivo con **becas específicas** (máster, doctorado, postgrado, formación continua, expertos, etc.)

# Estrategia Formativa en TL

Aitziber Atutxa, Núria Bel, Arantza Díaz de Ilarraza,  
Koldo Gojenola, Raquel Martínez, Patricio Martínez,  
Antoni Oliver, German Rigau, Alfonso Ureña