

NUEVOS ESPACIOS
PARA LAS LENGUAS INDÍGENAS:
Comunidades Urbanas, Medios de Comunicación y
Tecnologías de la Información

3. Tercer Conversatorio 9 de abril, 2014

Lenguas indígenas y nuevas tecnologías de la comunicación

¿Son compatibles las lenguas indígenas a las nuevas tecnologías de la comunicación?



Kepa Sarasola, representante del Grupo de Investigación Ixa (euskera)

1 Cronología de la recuperación del euskera

Escuela

Definición del euskera estandar

Universidad (Facultad de Informática)

Tecnología de la lengua, Grupo Ixa

2 Clasificación de las 7000 lenguas del planeta según su uso de tecnología

3 Primeros pasos para una lengua en tecnología

Nivel de una lengua con pocos recursos



PERÚ

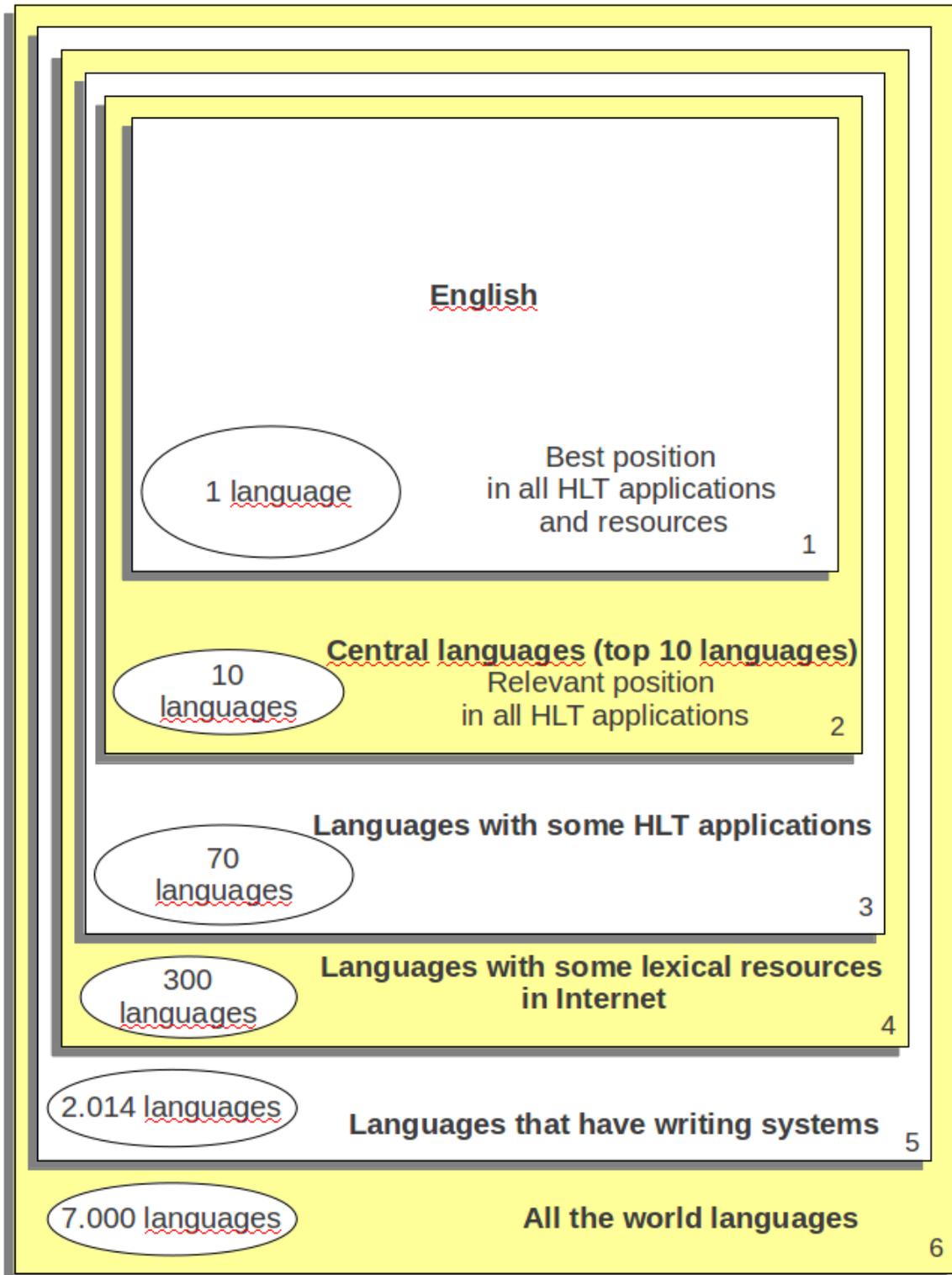
Ministerio de Cultura

Viceministerio
de Interculturalidad

“Año de la Promoción de la Industria Responsable y del Compromiso Climático”

Cronología del proceso de recuperación del euskera y tecnología de la lengua

Euskera en la enseñanza	Euskera estándar	Fac. de Informática UPV/EHU	Grupo Ixa	Quechua - Hinantin
1936-1960: el euskera está prohibido en las escuelas 1964: se crea la primera ikastola, escuela en la que todas las materias se dan en euskera.	1968: la Academia de la lengua vasca Euskaltzaindia fija las bases del euskera estándar (verbos, léxico básico y morfología)	1974: Se crea la Facultad de Informática de la Universidad del País Vasco (todo en español) 1984: En la Facultad de Informática de la Universidad del País Vasco se comienza a dar dos asignaturas en euskera.	1988: Se crea el grupo Ixa, que empieza a investigar como analizar automáticamente la palabras en euskera (muchos sufijos)	1985: alfabeto estandarizado
1983: La educación pública regula la creación de líneas en euskera (modelo D) y líneas en español y euskera (modelo B)		1992: Los tres primeros cursos de la Facultad de Informática de la Universidad del País Vasco se pueden estudiar en euskera	1994: El grupo Ixa presenta publicamente el corrector ortográfico para el euskera basado en el analizador morfológico y en el estándar definido por la Academia	
Modelos en euskera in crescendo	Léxico estándar in crescendo		1998: se crea la base datos léxica del euskera con 60.000 entradas 2001: lematizador, diccionarios electrónicos 2006: El grupo Ixa presenta el traductor de español a euskera MATXIN.	
Modelos en euskera in crescendo,	Léxico estándar in crescendo	1998: Los cinco cursos de la Facultad de Informática de la Universidad del País Vasco se pueden estudiar en euskera	2007-2013 mas de 20 productos de tecnología de la lengua para el euskera. Ixa no el único agente en este campo se crea un cluster para coordinar a más de 30 actores. 2014: Se hacen público dos recopilaciones de textos con 200 millones de palabras en euskera, para su consulta por Internet.	2012 analizador morfológico para quechua
2012: sólo un 15% de alumnos en escuelas sigue en el modelo A. (todo en español menos una asignatura sobre euskera)			2014: la base datos léxica del euskera contiene 160.000 entradas	2014: base de datos léxica, corrector ortográfico, sistemas de voz
Mas de 2/3 de los alumnos sigue el modelo D (todo en euskera menos una asignatura sobre español)				
Mas de la mitad de los estudiantes universitarios de la UPV/EHU estudian en euskera. En la práctica son trilingües (euskera, español, inglés)	El léxico estándar de la academia contiene 43.000 entradas			





DATOS GENERALES DE LA LENGUA									
NOMBRE DE LA LENGUA								Fecha	
CATEGORIA ASIGNADA A LA LENGUA	1	2	3	4	5	6	7	8	
POBLACIÓN									
DATOS GENERALES DEL ENCUESTADO									
NOMBRES Y APELLIDOS									
PROFESIÓN	1	2	3	4	5				
	Lingüística	Ciencias de la computación	Educación	Técnico	Otra profesión				

CAPÍTULO 200 - GRABACIONES / TEXTOS / DICCIONARIOS

- | | |
|--|---|
| <p>201 ¿Se desarrolla una recopilación de grabaciones orales?</p> <p>1 Si
2 No</p> <p>202 ¿Se desarrolla una recopilación de textos escritos (normalmente no estandarizados)?</p> <p>1 Si
2 No</p> <p>203 Listas de palabras, pequeños diccionarios.</p> <p>1 100 palabras
2 1.000 palabras
3 10.000 palabras
4 100.000 palabras</p> <p>204 Definición de un modelo unificado, estándar para la lengua.</p> <p>1 Si
2 No</p> | <p>205 Diccionario monolingüe para el modelo estándar. (versión accesible en Internet).</p> <p>1 1.000 palabras
2 10.000 palabras
3 100.000 palabras</p> <p>206 Diccionario bilingüe para el modelo estándar. (versión accesible en Internet)</p> <p>1 Si
2 No</p> <p>207 Corpus. Recopilación de textos escritos según el modelo estándar (versión accesible en Internet).</p> <p>1 10.000 palabras
2 100.000 palabras
3 > 1.000.000 palabras</p> |
|--|---|

CAPÍTULO 300 - GRAMÁTICA / APLICACIONES DE PROCESAMIENTO DEL LENGUAJE NATURAL

- | | |
|--|--|
| <p>301 Analizador morfológico (especialmente interesante para las lenguas con morfología rica como el euskara y el quechua).</p> <p>1 Si
2 No</p> <p>302 Wikipedia</p> <p>1 Si
2 No</p> <p>303 Reconocedor de textos escritos en esa lengua</p> <p>1 Si
2 No</p> <p>304 POS tagger / Lematizador</p> <p>1 Si
2 No</p> <p>305 Chunker</p> <p>1 Si
2 No</p> <p>306 Reconocedor de entidades.</p> <p>1 Si
2 No</p> <p>307 Gramática</p> <p>1 Si
2 No</p> <p>308 WorNet</p> <p>1 Si
2 No</p> | <p>309 Hay sistemas de tratamiento de voz?</p> <p>1 Si
2 No</p> <p>310 Otro tipo de productos</p> <p>1 Si
2 No</p> <p>311 Aplicaciones</p> <p>1 traducción automática.
2 reconocedor de voz,
3 síntesis de voz,
4 corrector ortográfico
5 ...</p> <p>312 ...</p> |
|--|--|

CAPÍTULO 400 - NOTAS

- *
* Establecidos esos niveles, podríamos fijar la posición (más o menos) de cada una de las lenguas de Perú.