

## Organización

- Invitados: viernes que viene Laura, Kadi Ferguson en abril
- Bibliografía:
  - MT an introductory guide- balkan (online)
  - *Computers and Translation: A translator's guide*, editor: H. Sommers 2003
  - *Diseño e implementación de sistemas de traducción automática*, J. Quesada, U. Sevilla, 2000
  - *Machine Translation: Its Scope and Limits*, Y. Wilks, Springer 2009
  - papers en mt-archive.info + algun libro de gigapedia

## Programa

- Traducción: herramientas y aplicaciones (TAO vs. TA)
- TA: historia, dificultades, aplicaciones
- Tipos de sistemas: basados en reglas, transferencia, estadísticos, basados en ejemplos
- Recursos necesarios
- Modelos de lenguaje
- TA estadística: modelos IBM, factorizados, por frases, decodificación
- Evaluación de la TA: métricas humanas, automáticas, basadas en la tarea y en el contexto
- Avances en el área de la TA

## Qué es la TA?

- Según IAMT: "Es la aplicación de computadoras a la tarea de traducir textos de un lenguaje natural a otro"
- Según Systran: "*Machine translation (MT) is automated translation. To process any translation, human or automated, the meaning of a text in the original (source) language must be fully restored in the target language [...]*"
- También abarca la traducción de habla, lenguaje de señas, Braille, ...

## Por qué hacemos TA?

- Como individuos: para comunicarnos en idiomas que no necesariamente conocemos
- Como empresas: para mantener o incrementar ganancias y/o mercados
- Para acelerar proceso de traducción, procesar mayor volumen de textos
- etc etc ...

## Cuánto se usa la TA?

- En 2000 *BabelFish* de Altavista (iniciado en 1997) se usaba 1 millón de veces por día
- En 2003 portales como Altavista y Google procesan cerca de 10 millones de pedidos idiomas mayoritarios
- En 2009 Google cuenta con 51 pares de lenguajes y herramientas para post-editar
- Multinacionales con demandas multilingües: ej. Symantec usa Systran para traducir a 27+ idiomas. Muchas otras Cisco, Caterpillar, Ford, Chrysler, ...

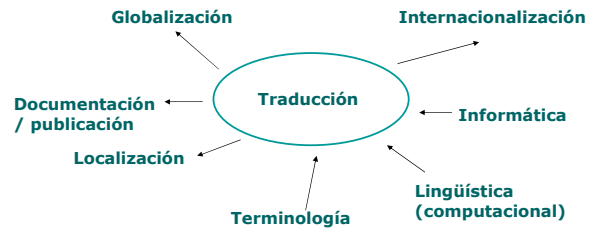
## Cuánto más puede usarse?

- Volumen de textos a traducir excede la capacidad de traductores
- Crecimiento exponencial de información disponible (blogs, redes sociales, diarios, etc)
- Automatizar para cumplir expectativas de usuarios, managers, compradores, ...

## Cuánto más podría usarse?

- Durante la expansión de empresas (globalización más adelante)
- En contextos como
  - Inclusión de nuevos miembros en Unión Europea
  - Situaciones excepcionales que generan necesidades urgentes (ej. terremoto en Haití textos en Creole)
  - Para preservar lenguas en peligro (Amerindias, dialectos, etc)
  - Para detectar copias/plagios en patentes, papers en conferencias, ...
  - En documentación automática (tipo javadoc)
  - Muchas otras situaciones que no se me ocurrieron!!

## Importancia de la traducción



## Importancia de la traducción

- **Globalización:** Actividades realizadas con el propósito de comercializar un producto o servicio en mercados regionales
- **Internacionalización (I18N):** Desarrollar un producto (software) de tal manera que sea sencillo adaptarlo a otros idiomas y culturas
  - Este proceso se lleva a cabo durante la fase de diseño y desarrollo del producto.
  - Evitar referencias culturales, y aislar elementos que sean específicos del idioma (formato de fechas, números, moneda, etc.)

## Importancia de la traducción

- **Localización (L10N):** Adaptación de un producto o servicio a un mercado regional (Localidad - *locale*)
  - Es llevada a cabo por traductores, localizadores, ingenieros
  - Industria en crecimiento en Argentina
    - Erikssen Cba, Ushuaia / lingua graphics /connections Rosario, etc
- **Localidad:** características del entorno del usuario que dependen del lenguaje, país, región y de convenciones culturales (ej. ordenación, diseño del teclado, fecha, hora, números y formato de moneda)

## Niveles de intervención humana

- Human translation (HT)
  - Interpretación del significado de un texto y la producción *libre* de un equivalente en otro idioma
- Human-aided automatic (HAMT)
  - Humanos intervienen en el proceso de TA para disambiguar sentidos, elegir vocabulario, etc
- Machine-aided human translation (MAHT)
  - Uso de herramientas para ayudar al traductor humano – manejo terminología, memorias de traducción, etc
- Fully automatic (MT) = sin intervención

## Áreas involucradas

- Lingüística
  - Gramática, morfología, sintaxis de un idioma
- Traducción y/o interpretación
  - Comprensión de idiomas involucrados (source/target)
- Lingüística computacional
  - Procesamiento automático de problemas lingüísticos
- Ciencias de la computación
  - Implementación de sistemas, herramientas de soporte
- Matemáticas
  - TA estadística, optimización de heurísticas, etc

## Ejemplo: oferta trabajo en TA

- The successful applicant must have the following qualifications:
  - Native or near-native fluency in French
  - M.S. in Computer/Information Science (or related field), or equivalent experience
  - Strong programming skills in Lisp and Unix
  - Experience communicating and working productively within a development group
  - Experience in designing and implementing multi-module software systems
- The following skills are also desirable:
  - Coursework in linguistics or computational linguistics
  - Experience in developing machine translation software or other types of language processing software
  - Experience with commercial software development and system engineering
  - Fluency in German, Spanish, Portuguese, Russian, Italian
  - Some familiarity with PERL, C++

## Qué/Dónde se puede hacer?

- Trabajo final, doctorado, pasantías, trabajo en la industria, ...
- Arg.: grupo PLN FaMAF, UBA, empresas como Erikssen
  - En famaf análisis sintáctico, evaluación de la TA, TA, sistemas de diálogos, minería de dato
- Afuera: ver listas mt-list, linguist-list, 7FP EU, EUA (microsoft, univesidades)