

Introducción a la Traducción Automática

CURSO 2008-2009

Mikel L. Forcada y Juan Antonio Pérez-Ortiz

{mlf, japerez}@dlsi.ua.es

Departamento de Lenguajes y Sistemas Informáticos

Objetivos de los cursos

Estos son los objetivos que se quiere conseguir al final del curso, algunos de los cuales pueden haber sido conseguidos por parte del alumnado antes de cursarlo.

- Que sean conscientes de las aplicaciones posibles de la traducción automática y de los problemas que comporta.
- Que conozcan los conceptos lingüísticos básicos necesarios para discutir los problemas y las soluciones: morfología (categorías léxicas más comunes, análisis morfológico, homografía), sintaxis (sintaxis generativa básica, árboles de análisis sintáctico, adjunción, movimiento, etc.), semántica léxica (polisemia, hiponimia, hiperonimia, unidades multipalabra, etc.) y oracional (principio de composicionalidad semántica).
- Que conozcan los conceptos informáticos básicos sobre los cuales se basan las soluciones: algoritmos y programas, representación de los textos en el ordenador (codificación, formados, estructuración de los documentos, XML, etc.).
- Que comprendan las operaciones y las estructuras lingüístico-computacionales básicas sobre las cuales se basan los sistemas de traducción automática: división del texto en palabras, diccionarios morfológicos, diccionarios bilingües, analizadores morfológicos, gramáticas, analizadores sintácticos, etc.
- Que conozcan el estado actual de la cuestión y las soluciones más exitosas a la traducción automática: sistemas de transferencia; sistemas de *interlingua*; sistemas basados en corpus (estadísticos, basados en ejemplos) e híbridos.
- Que sean conscientes de la complejidad intrínseca de la traducción automática y de las causas de esta complejidad.

Resumen del programa

Los bloques básicos del curso son:

1. Introducción
 - ¿Qué es (y que no es) la traducción automática (TA)?
2. Aplicaciones de la traducción automática.
 - Interés, aplicaciones y usos de la TA.
 - Es posible la TA perfecta de propósito general?
 - Grados de dificultad de las tareas de traducción automática.
 - Aplicabilidad de la TA real (imperfecta):
 - asimilación y diseminación
 - *human-aided machine translation* y *machine-aided human translation*
 - preedición y postedición
 - lenguajes controlados

3. Ambigüedad y problemas relacionados.
 - El principio de composicionalidad semántica.
 - Ambigüedad léxica
 - Ambigüedad sintáctica.
 - Ambigüedades mixtas.
 - Otras fuentes de ambigüedad.
 - Otros problemas no relacionados con el análisis del texto origen (p.e., generación).
4. Técnicas de traducción automática (programación deductiva)
 - Problemas de la traducción palabra a palabra.
 - Traducción indirecta: el modelo de transferencia..
 - Traducción indirecta por transferencia morfológica (procesamiento de patrones).
 - Análisis morfológico. Diccionarios morfológicos. Problemática asociada.
 - La traducción indirecta por transferencia sintáctica.
 - Formalismos gramaticales: gramáticas independientes del contexto. Árboles. Analizadores sintácticos ascendentes y descendentes.
 - Problemas de la transferencia sintáctica.
 - La traducción por transferencia semántica.
 - Hacia la neutralidad lingüística: los sistemas de interlingua.
 - Ventajas de los sistemas basados en transferencia.
 - Análisis de las interlinguas.
 - La traducción automática real: entre la aplicación ideal del principio de composicionalidad semántica y el modelo palabra por palabra.
5. Programación (totalmente o parcialmente inductiva) de traductores automáticos.
 - Memorias de traducción.
 - Sistemas basados en ejemplos.
 - Sistemas estadísticos.
 - Inferencia de traductores sencillos.
6. Evaluación de la traducción automática.
 - Aspectos básicos de la evaluación de los sistemas de traducción automática.
 - Sobre la comparación de los traductores automáticos con los traductores humanos.

Mecánica de las clases

Además de las sesiones en el aula, que se procurará que sean lo más participativas posible, en cada sesión de laboratorio se propondrá una actividad concreta (de análisis o de modificación de los datos lingüísticos de software existente) que se tendrá que resolver individualmente (de manera más o menos dirigida).

Las sesiones de laboratorio se irán anunciando con antelación suficiente a lo largo del curso.

Es especialmente importante que, después de cada clase, cada estudiante dedique un tiempo a hacer un esfuerzo para sintetizar la discusión de clase, las explicaciones del profesorado, y la bibliografía recomendada. Es cuando se hace esta síntesis cuando es posible que se planteen dudas y problemas que se deban discutir con el profesorado en las horas de asistencia al alumnado (anunciadas en el web del Departamento de Lenguajes y Sistemas Informáticos)

De hecho, en algunas sesiones se encargará que el alumnado lea algún artículo concreto sobre traducción automática y que lo resuma en una presentación para el resto de la clase.

Evaluación

Para superar el curso se debe asistir regularmente a las sesiones de aula y de laboratorio e involucrarse activamente en su desarrollo y presentar los trabajos que se encarguen, que incluyen dos prácticas de laboratorio con sistemas reales de traducción automática y la presentación (con transparencias) del resumen de un artículo asignado por el profesorado. La participación del alumnado en las actividades influirá en su calificación.

Bibliografía

- Alcaraz Varó, E. and Martínez Linares, M. (1997). Diccionario de lingüística moderna. Ariel, Barcelona.
- Arnold, D. (1993). Sur la conception du transfert. In Bouillon, P. and Clas, A., editors, *La traductique*, pages 64-76. Presses Univ. Montréal, Montréal. English translation available at <http://clwww.essex.ac.uk/~doug/papers/transfer.ps.gz>.
- Arnold, D. (2003). Why translation is difficult for computers. In Somers, H., editor, *Computers and Translation: A translator's guide*. Benjamins, Amsterdam/Philadelphia. Disponible en <http://www.essex.ac.uk/linguistics/research/clgroup/papers/doug/TranslatBook.ps>.
- Arnold, D., Balkan, L., Meijer, S., Humphreys, R., and Sadler, L. (1994). *Machine Translation: An Introductory Guide*. NCC Blackwell, Oxford. Disponible en <http://clwww.essex.ac.uk/~doug/MTbook/>.
- Forcada, M. L. and Pérez Ortiz, J. A. (2007). *Informàtica Aplicada a la Traducció: notes de classe amb problemes i exercicis resolts*. Universitat d'Alacant. Disponible en <http://www.dlsi.ua.es/~mlf/iat/l libreiat/>.
- Hutchins, W. and Somers, H. (1992). *An introduction to machine translation*. Academic Press. (hay traducción española, *Introducción a la traducción automática*, editada por Visor en 1995).
- Ide, N. and Veronis, J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24:1-40.
- [Jacqmin 1993]
- Jacqmin, L. (1993). Classification générale des systèmes de traduction automatique. In P., B. and Clas, A., editors, *La traductique*. Presses Univ. Montréal, Montréal.
- Lyovin, A. (1997). *Languages of the world*. Oxford Univ. Press, Oxford.
- Matthews, P. (1997). *The concise Oxford dictionary of linguistics*. Oxford Univ. Press, Oxford.
- Mohri, M. (1997). Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(2):269-311.
- Radford, A., Atkinson, M., Britain, D., Clahsen, H., and Spencer, A. (1999). *Linguistics: an introduction*. Cambridge Univ. Press, Cambridge.
- Ramos, J. R. (1992). *Introducció a la sintaxi*. Tàndem, València.
- Roche, E. and Schabes, Y., editors (1997). *Finite-State Language Processing*. MIT Press, Cambridge, Massachusetts.
- Sager, J. C. (1993). *Language engineering and translation: consequences of automation*. Benjamins, Amsterdam
- Somers, H., editor (2003). *Computers and Translation: a Translator's Guide*. John Benjamins,

Amsterdam/Philadelphia.

Tuson, J. (1999). Com és que ens entenem? (si és que ens entenem). Empúries, Barcelona.

Wong, S. (2003). Machine translation techniques. Disponible en
<http://www.fi.muni.cz/usr/wong/teaching/mt/notes.ps>.

Zwarts, J., Ackema, P., Elenbaas, N., Esther Janse, Iris Mulders, J. R., Visch, E., van Gestel, M., van der Laan, B., and Leidekker, H. (2003). Lexicon of linguistics. <http://tristram.let.uu.nl/UiL-OTS/Lexicon/>.