

A 45-hour Computers in Translation course

Mikel L. Forcada

Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant, E-03071 Alacant (Spain)

T⁴ workshop, MT Summit IX, New Orleans 2003

Index

Index

- The subject

Index

- The subject
- Students, groups and sessions

Index

- The subject
- Students, groups and sessions
- Methodology

Index

- The subject
- Students, groups and sessions
- Methodology
- Syllabus

Index

- The subject
- Students, groups and sessions
- Methodology
- Syllabus
- Bibliography

Index

- The subject
- Students, groups and sessions
- Methodology
- Syllabus
- Bibliography
- Closing comments

The subject

The subject

Informática Aplicada a la Traducción (official Spanish name)

The subject

Informática Aplicada a la Traducción (official Spanish name)

Mandatory for 4- or 5-year translation degrees in Spain

The subject

Informática Aplicada a la Traducción (official Spanish name)

Mandatory for 4- or 5-year translation degrees in Spain

Minimum of 4.5 credits (=45 h)

The subject

Informática Aplicada a la Traducción (official Spanish name)

Mandatory for 4- or 5-year translation degrees in Spain

Minimum of 4.5 credits (=45 h)

Official description (quite short and open to interpretation):

Access to the necessary tools for translation work. Machine translation and computer-assisted translation. System Integration.

The subject

Informática Aplicada a la Traducción (official Spanish name)

Mandatory for 4- or 5-year translation degrees in Spain

Minimum of 4.5 credits (=45 h)

Official description (quite short and open to interpretation):

Access to the necessary tools for translation work. Machine translation and computer-assisted translation. System Integration.

Subject expected to provide future translators with **all** they need to know about computers in translation (!).

Students, groups and sessions

University of Alacant:

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.
- Two 75-student classroom groups

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.
- Two 75-student classroom groups
- Six 25-student laboratory groups

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.
- Two 75-student classroom groups
- Six 25-student laboratory groups
- 45 hours (no extension beyond official minimum).

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.
- Two 75-student classroom groups
- Six 25-student laboratory groups
- 45 hours (no extension beyond official minimum).
- 30 1.5-hour sessions (19 classroom, 11 laboratory)

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.
- Two 75-student classroom groups
- Six 25-student laboratory groups
- 45 hours (no extension beyond official minimum).
- 30 1.5-hour sessions (19 classroom, 11 laboratory)
- 6 office hours a week per instructor (presential interaction)

Students, groups and sessions

University of Alacant:

- 150 students: German 30, English 60, French 60.
- Two 75-student classroom groups
- Six 25-student laboratory groups
- 45 hours (no extension beyond official minimum).
- 30 1.5-hour sessions (19 classroom, 11 laboratory)
- 6 office hours a week per instructor (presential interaction)
- virtual campus (nonpresential interaction)

Methodology: Classroom work /1

Methodology: Classroom work /1

Classroom work organized around an activity program (sequence of activities)

Methodology: Classroom work /1

Classroom work organized around an activity program (sequence of activities)

Activities pose open problems **before** the theory is explained.

Methodology: Classroom work /1

Classroom work organized around an activity program (sequence of activities)

Activities pose open problems **before** the theory is explained.

Example:

Ambiguity is an essential feature of natural languages. Could you write up a formal definition of ambiguity? Why do you think human language is ambiguous? Why does ambiguity make machine translation difficult?

(followed by an activity where students have to devise a linguistically motivated classification from a set of ambiguous sentences).

Methodology: Classroom work /2

Methodology: Classroom work /2

- Activities are first tackled individually,

Methodology: Classroom work /2

- Activities are first tackled individually,
- then discussed in 3-student groups (ideally stable groups),

Methodology: Classroom work /2

- Activities are first tackled individually,
- then discussed in 3-student groups (ideally stable groups),
- and are finally discussed by the whole classroom group

Methodology: Classroom work /2

- Activities are first tackled individually,
- then discussed in 3-student groups (ideally stable groups),
- and are finally discussed by the whole classroom group
- The instructor integrates the discussion in a “classical” lecture.

Methodology: Classroom work /2

- Activities are first tackled individually,
- then discussed in 3-student groups (ideally stable groups),
- and are finally discussed by the whole classroom group
- The instructor integrates the discussion in a “classical” lecture.

Example: explains a classification of ambiguity based on the principle of compositional semantics.

Methodology: Classroom work /3

Methodology: Classroom work /3

By working in this way, students

Methodology: Classroom work /3

By working in this way, students

- analyse the problem,

Methodology: Classroom work /3

By working in this way, students

- analyse the problem,
- may even advance part of the solution, but at least

Methodology: Classroom work /3

By working in this way, students

- analyse the problem,
- may even advance part of the solution, but at least
- get ready to understand the instructor's explanation of the solution.

Methodology: Classroom work /3

By working in this way, students

- analyse the problem,
- may even advance part of the solution, but at least
- get ready to understand the instructor's explanation of the solution.

and instructors

Methodology: Classroom work /3

By working in this way, students

- analyse the problem,
- may even advance part of the solution, but at least
- get ready to understand the instructor's explanation of the solution.

and instructors

- learn what students already know about the problem

Methodology: Classroom work /3

By working in this way, students

- analyse the problem,
- may even advance part of the solution, but at least
- get ready to understand the instructor's explanation of the solution.

and instructors

- learn what students already know about the problem
- use this knowledge to anchor the explanation of new, complex concepts.

Methodology: Classroom work /4

Methodology: Classroom work /4

After the classroom:

Methodology: Classroom work /4

After the classroom:

- students are expected to make a synthesis

Methodology: Classroom work /4

After the classroom:

- students are expected to make a synthesis
- and apply acquired knowledge to new problems

Methodology: Classroom work /4

After the classroom:

- students are expected to make a synthesis
- and apply acquired knowledge to new problems

During office hours, teachers help the synthesis by clearing doubts and providing guidance.

Methodology: Laboratory work/1

Methodology: Laboratory work/1

One assignment per session (some assignments take two sessions).

Methodology: Laboratory work/1

One assignment per session (some assignments take two sessions).

Students work individually or in pairs.

Methodology: Laboratory work/1

One assignment per session (some assignments take two sessions).

Students work individually or in pairs.

Example #1 (lab session L_6): students analyse what a given commercial MT system does besides simply substituting words:

Methodology: Laboratory work/1

One assignment per session (some assignments take two sessions).

Students work individually or in pairs.

Example #1 (lab session L_6): students analyse what a given commercial MT system does besides simply substituting words:

- first make it translate words in isolation

Methodology: Laboratory work/1

One assignment per session (some assignments take two sessions).

Students work individually or in pairs.

Example #1 (lab session L_6): students analyse what a given commercial MT system does besides simply substituting words:

- first make it translate words in isolation
- then in sentences

Methodology: Laboratory work/1

One assignment per session (some assignments take two sessions).

Students work individually or in pairs.

Example #1 (lab session L_6): students analyse what a given commercial MT system does besides simply substituting words:

- first make it translate words in isolation
- then in sentences

and study the differences (Perez-Ortiz and Forcada, TMT 2001).

Methodology: Laboratory work/2

Example #2 (lab session L_7): students run a set of instructor-designed increasingly-complex noun phrases through an MT system to infer its word reordering rules (Forcada, MT 2000).

Syllabus/1

Syllabus/1

Syllabus design started in 1995, before LETRAC (Badia et al. 1999) or Balkan et al.'s survey (1997).

Syllabus/1

Syllabus design started in 1995, before LETRAC (Badia et al. 1999) or Balkan et al.'s survey (1997).

Evolved into 10 blocks ($B_1 \dots B_{10}$).

Syllabus/1

Syllabus design started in 1995, before LETRAC (Badia et al. 1999) or Balkan et al.'s survey (1997).

Evolved into 10 blocks ($B_1 \dots B_{10}$).

Blocks contain classroom ($C_1 - C_{19}$) and laboratory ($L_1 - L_{11}$) sessions.

Syllabus/2

Block: B_1 : What are we going to study?

Objective: Knowing how computers may be applied to translation: automatable and non-automatable translation tasks; machine translation; human-aided machine translation; machine-aided human translation.

Classroom sessions: C_1 (week 1).

Lab sessions: None.

Syllabus/3

Block B_2 : Computers and programs

Objective: Acquiring basic personal computer concepts: hardware and software; memory and storage; files and directory structure; computer programs; CPUs; operating systems.

Classroom sessions: $C_2 - C_4$ (weeks 2 and 3).

Lab sessions: L_1 (week 3: analysing the hardware characteristics of the PC in the lab; creating and modifying a directory structure on a diskette).

Syllabus/4

Block B_3 : Internet basics

Objective: Acquiring basic concepts about the internet and about its application to the translation task: networks, the internet, services (lexical databases, dictionaries, encyclopedia, texts, bitexts, search engines), addressing; home access.

Classroom sessions: C_5 (week 3).

Lab sessions: L_2 and L_3 (weeks 4 and 5: searching for translations with Google; basics of HTML; building a webpage from a template and publishing it).

Syllabus/5

Block B_4 : Texts and formats

Objective: Learn basic concepts about the storage, format, structuring, presentation, creation and manipulation of text documents: character encoding; formats for presentation and structuring of content; XML; stylesheets; OCR and speech recognition.

Classroom sessions: C_6 and C_7 (weeks 4 and 5).

Lab sessions: L_4 and L_5 (weeks 6 and 7: validating XML documents against a simple DTD; tagging a text according to a certain DTD and validating it).

Syllabus/6

Block B_5 : Machine translation and applications

Objective: Learning how real machine translation is applied in the real world despite its imperfections: assimilation and dissemination; human-aided machine translation (preediting, postediting, interaction, controlled languages); MT as a component of communication systems; nonlinguistic requirements (speed, format preservation).

Classroom sessions: C_8 (week 6).

Lab sessions: none.

Syllabus/7

Block B_6 : Ambiguity

Objective: Identifying ambiguity as the main source of errors in machine translation: lexical, structural, and mixed ambiguity; ambiguity resolution in MT systems.

Classroom sessions: C_9 and C_{10} (weeks 7 and 8).

Lab sessions: none.

Syllabus/8

Block B_7 : How does machine translation work?

Objective: Knowing the main machine translation strategies and their implementation as distinct, consecutive phases or tasks: commercial systems as intuitive refinements over word substitution; transfer; interlingua; inductive strategies.

Classroom sessions: $C_{11} - C_{14}$ (weeks 9 and 10).

Lab sessions: L_6 and L_7 (weeks 8 and 11: “machine translation is not word by word”, and “discovering reordering and agreement rules”).

Syllabus/9

Block B_8 : Machine translation evaluation

Objective: Learning to use knowledge about how MT systems work to evaluate them with an adequate technical level and well-founded criteria: identifying aspects to be evaluated and their relative importance; recognizing the inadequacy of comparison with human translation.

Classroom sessions: C_{15} (week 11).

Lab sessions: L_8 (week 12: evaluation and classification of MT errors in real texts).

Syllabus/10

Block B_9 : Lexical databases

Objective: Learning basic concepts about databases: tables, records, fields, ordering and indexing for faster search; using concept-oriented lexical databases for specialized translation and terminological coherence. Being able to design, create and maintain a lexical database using suitable software.

Classroom sessions: C_{16} (week 12).

Lab sessions: L_9 (week 13: creating a small lexical database and performing searches over it).

Syllabus/11

Block B_{10} : Translation memory

Objective: Understanding the importance of translation memories (TM) as an efficient solution to human translation with a high degree of repetitiveness: TMs as databases of translation units; bitext processing; pre-translation; advantages of TM-based translation work; TMX for interchange.

Classroom sessions: $C_{17} - C_{18}$ (weeks 13 and 14).

Lab sessions: L_{10} (week 14: a taste of the complete TM cycle: alignment of a bitext followed by pre-translation and correction of a new text and TM updating).

Syllabus/12

Comparing with LETRAC:

- 45 h much less than 230 h recommended by LETRAC: sacrifices!
- MT mandatory in Spain but optional in LETRAC.
- No desktop publishing in Alacant.
- No XML in LETRAC (too early: XML 1998).
- Little terminology in Alacant (but other mandatory subjects relating terminology have 10.5 credits).

Bibliography

In addition to yearly-updated class notes in Catalan (PDF),

- The classical Hutchins and Somers' (1992) book
- Arnold et al.'s (1994) online book.
- Trujillo's (1999) Translation engines
- [new!] Somers, H., ed. (2003) Computers in Translation: A translator's guide
- Bouillon and Clas, eds. (1993) La Traductique.
- Cole, ed. (1996) Survey of the State of the Art in HLT

and many others... (see paper).

Closing comments

Closing comments

It's hard to decide what to teach in 45 hours.

Closing comments

It's hard to decide what to teach in 45 hours.

It's hard to describe a solution in 6 pages (or 25 slides)!

Closing comments

It's hard to decide what to teach in 45 hours.

It's hard to describe a solution in 6 pages (or 25 slides)!

But...

Closing comments

It's hard to decide what to teach in 45 hours.

It's hard to describe a solution in 6 pages (or 25 slides)!

But...

- if you have similar restrictions

Closing comments

It's hard to decide what to teach in 45 hours.

It's hard to describe a solution in 6 pages (or 25 slides)!

But...

- if you have similar restrictions
- or teach in a similar environment

Closing comments

It's hard to decide what to teach in 45 hours.

It's hard to describe a solution in 6 pages (or 25 slides)!

But...

- if you have similar restrictions
- or teach in a similar environment

and you think this proposal helps you...

Closing comments

It's hard to decide what to teach in 45 hours.

It's hard to describe a solution in 6 pages (or 25 slides)!

But...

- if you have similar restrictions
- or teach in a similar environment

and you think this proposal helps you...

...I'll be happy to talk to you and even translate some of my materials.