So you want to be a linguist?

L2 LEA PARCOURS “INDUSTRIES DE LA LANGUE”

CLASS 1 : MÉTIERS DE LA LANGUE / WORKING WITH LANGUAGE
Aims

➢ What does it mean to **work in the language industry?**
  
  *process-oriented / product-oriented* -> **linguist-oriented focus**

➢ Overview of sectors and jobs in the language industry:
  ➢ What **skills** are needed?
  ➢ How can you acquire and perfect these skills?
  ➢ How can you make a living?
  ➢ How is the sector evolving?

➢ The importance of **process, specialisation** and **quality control**

➢ Digital tools and digital environments: working with (not against) machines.
Assessment

➢ This class is technically a lecture (CM), but evaluation is by continuous assessment (CC). You are required to submit two assignments:

1. A group assignment, to begin in class on 02/03 and submit by 10/03
2. An individual research project, instructions distributed in March, due 04/05.

➢ Attendance in class is compulsory. Participation forms part of your grade.

➢ Your research project will require doing some research/reading outside class. You'll find a selection of resources in our online bibliography, and lots more in the library catalogue.

➢ Assessments can be written in French or English. If you write in your native language, all language errors will be penalised!
Plubel

➢ Slides, links, worksheets and other resources will be shared here
➢ https://moodle-campus.u-bourgogne.fr/course/view.php?id=4781
➢ Guest access possible with password LEA2
<table>
<thead>
<tr>
<th>No.</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>03/02</td>
<td>Introduction: So you want to be a linguist?</td>
</tr>
<tr>
<td>2</td>
<td>10/02</td>
<td>Translation processes and environments</td>
</tr>
<tr>
<td>3</td>
<td>17/02</td>
<td>Translation and CAT tools; localising a mobile app (Holidays)</td>
</tr>
<tr>
<td>4</td>
<td>02/03</td>
<td>Technical writing (start Assignment 1)</td>
</tr>
<tr>
<td>5</td>
<td>09/03</td>
<td>Introduction to interpreting (Assignment 1 due)</td>
</tr>
<tr>
<td>6</td>
<td>16/03</td>
<td>Interpreting: from theory to practice</td>
</tr>
<tr>
<td>7</td>
<td>06/04</td>
<td>Audiovisual translation: introduction to revoicing</td>
</tr>
<tr>
<td>8</td>
<td>18/03</td>
<td>Accessibility in practice (deadline to confirm research topic)</td>
</tr>
</tbody>
</table>

(Holidays)

(Assignment 2 deadline 04/05)
<table>
<thead>
<tr>
<th>Parcours « Stratégie de marques et d’entreprises »</th>
<th>Parcours « Industries de la langue »</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication et marketing (CM+TD)</td>
<td>Métiers de la langue</td>
</tr>
<tr>
<td>Comportement du consommateur (TD)</td>
<td>Translation studies et linguistique appliquée</td>
</tr>
<tr>
<td></td>
<td>Langue et relations internationales</td>
</tr>
</tbody>
</table>
the language industry in all its diversity
What is a linguist/language professional?

1. Someone who **uses language as the main tool of their trade** (cf. language industry diagram)

2. Someone engaged in **linguistics**, or the scientific study of language.

- These definitions are not contradictory: research and practice often overlap, especially in new/developing fields!
  - “Applied Linguistics addresses a broad range of real-life language-related issues by developing its own knowledge-base about language, its users and uses, and their underlying social and material conditions.” ([https://www.aaal.org](https://www.aaal.org))

- Both these definitions go far beyond the ability to learn or communicate in a foreign language!

- A professional linguist may work in fields like tourism, teaching or international trade – but language remains their main focus!
Challenges for language professionals

- What challenges do the following represent for the language industry?
  - Globalisation of products
  - Globalisation of workers and working conditions
  - Growth of digital products
  - Growth of digital tools
  - Changes in user expectations
  - Accessibility
  - Others?

- How can YOU address these?
Professionalism and quality standards

- Trained language professionals cannot and should not survive by aiming for the bottom of the market...
  - because someone else will always work for less
  - because basic tasks are increasingly automated.
- Language professionals should always aim to produce quality work... and charge accordingly.
- Language quality is not a sufficient criterion but is still necessary!
- Where standards and guidelines exist, follow them!
- Educating your clients about quality benefits everyone!
Case study: understanding machine translation

➢ As a future professional linguist, what do you think of machine translation, and other forms of natural language processing?

➢ Does it worry you? Should it?

➢ Since it’s not going away, what can you do about it?

❑ Tip 1: learn how to work with digital tools, not against them!
  ❑ More on CAT environments and MT postediting next week.

❑ Tip 2: think carefully about the implications of ‘free’ tools!
  ❑ Your client’s image is at stake – this is why a smart client relies on professionals!
  ❑ Your client is king, but may not always be right – and it’s a professional’s job to tell them.
  ❑ “If you’re not paying, then you’re the product”: Google Translate mainly exists in order to benefit Google!
Types of machine translation

- **Rule-based** (earliest approaches, still used in some contexts)
- **Statistical** (most widespread approach today)
- **Neuronal** (emerging approach using AI/Deep Learning)

➢ *In practice, many tools use aspects of all three approaches.*

➢ *MT is not quite at the stage of replacing human intelligence:*

   “A translator must be at ease with the two languages involved, and he must have the special skills to reformulate a source language in a target language that does not have the same wording or the same structure. These kinds of skills are not directly available to machines. [...] To be able to reformulate a sentence, one must of course have a good command of the language itself, but one must also master the search for analogy between concepts, which is more complicated than just equivalences between words and expressions.” (Poibeau 2017: 12)
Statistical machine translation (Google vs. Jane Austen)

- TL sentence comprehensible but not perfect: ‘il est une vérité’, ‘dans le besoin d’une femme’
  - Phrase-based analysis (it is <NP/AdjP> that <NP/SUBJ> must be in want of <NP>)
  - Concatenation of existing phrases, sometimes entire sentences, using web as corpus.
- Ambiguity: “single” in SL sentence translated with more common (but incorrect) interpretation: seul rather than célibataire, even though a human can easily recognise the context.
- But algorithms are constantly improving, especially when supplied with large amounts of data and feedback.
Statistical machine translation (Google vs. Tolstoy)

Here, statistical analysis seems to have located the sentence, or sentences with very similar construction, in existing translations. Try searching for the French result using Google...
Statistical machine translation
(Google vs. Tolstoy with pivot language)
Rule-based machine translation (Systran vs. Tolstoy)

• Rule-based approaches work well:
  • For language pairs with deep lexical and syntactical similarities (e.g. French<>Catalan)
  • In specialised contexts using constrained language/lexical sets.
    • Cf. early MT research on Cold War nuclear physics (EN<>RU) and on Swiss avalanche warnings (DE<>FR<>IT)
  • In this example: we see a notably more ‘literal’ translation than in the previous slide.
Rule-based machine translation (Systran vs. Tolstoy with pivot language)

- Relay language dissimilar to SL or TL: note badly-formed TL sentence.
- Polysemy: words (or, in logographic languages such as Chinese, written characters) can have more than one meaning!
Syntactic patterns: Systran’s engine expects the more frequent construction “All <NP> are <NP>” rather than “All <NP> are <AdjP>”, leading to “All happy families are [the] elephants”.

The relative frequency of the two results in Chinese may also explain the result: one meaning is more common than the other, even though it’s inappropriate in this context!

Rule-based translation engines sometimes also use statistics!
Machine translation + voice recognition (Microsoft Translator)

- Single, simple sentence
- SVO structure: same in SL and TL
- Distinct sounds, little or no ambiguity in pronunciation or intonation
- Possible lexical (and cultural) ambiguity (beyond the scope of MT engines):

I love bacon

J'adore le bacon
“It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife.”

(Jane Austen)
Voice recognition issues:
- Pause between sentences
- Misrecognition of sewer as *sua*: pronounced /suːə/ in non-rhotic Aus or UK English vs. /suːr/ or /suːɚ/ in rhotic General American English.
- Misrecognition of *what you put in* (phrasal verb with stress on postpositional *in*) as *what you mean* (statistically more probable)

Machine translation issues:
- Reasonably well-formed translation of transcribed source text, discounting unconjugated verb sortir. Note choice of pronoun *vous*. 
- We might have expected an issue translating *a sewer > les égouts*, had the software recognised the SL expression!
- Problem dealing with context: no likely statistical model here!
- Humour depends on ambiguity.
  This joke is translatable, but determining that this is the case and deciding how to go about it requires creative (human) intelligence.

‘Life is like a sewer. What you get out of it depends on what you put in.’
(Tom Lehrer)
Machine translation with text recognition. (Microsoft)

- OCR issues (lighting, camera resolution, angle, moving object)
- Some lexical items recognised (CPU, RAM, USB, Getting started>Mise en route)
  ➢ Does the product domain (IT) facilitate statistical analysis?
- Traduction de noms propres:
  EN: <logo rond> Raspberry Pi
  FR: O Pi Framboise
Machine translation and specialised domains – what’s the problem here?
Activity

- Working in pairs or small groups, locate your favourite machine translation app or website.
- Find a language combination and source text that produces results that seem incoherent or amusing.
- Try to analyse why the results are incoherent/amusing – the preceding slides should give you some possible angles to try.
For next week:

Please (re-)read the following resources, and come prepared to discuss:

- Daniel Gouadec’s discussion of translation as a professional activity at http://www.profession-traducteur.net/traduction/traduction.htm
- Translation, getting it right/Traduction, faire les bons choix
- Translation, buying a non-commodity/Les mots au kilo
- (both available from https://www.sft.fr/publications.html)

Find out what you can about the following question: *What is computer-assisted translation (CAT) and how is it different from machine translation?*

Take a look at the CAT tools [WordFastAnywhere](https://www.sft.fr/publications.html) (online, create a free account) and/or [OmegaT](https://www.sft.fr/publications.html) (free to download and install). We will be using these in class and for your first group assignment.