

# 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**

## L2 LEA - parcours "Industries de la langue"

[Dashboard](#) ▶ [Courses](#) ▶ [UFR](#) ▶ [UFR Langues et Communication](#) ▶ [05LLEA - Licence langues étrangères appliquées](#) ▶ [Cours Multi Licences](#) ▶ [L2 LEA - parcours IdL](#)

 [Forum des nouvelles](#)

### Semestre 3 : Introduction aux industries de la langue

Diapos, liens, et autres ressources utiles pour le cours du semestre 3.

*Consultez cet espace régulièrement : il n'y aura pas de distribution de photocopies !*

#### SEARCH FORUMS

Go

[Advanced search](#) 

#### LATEST ANNOUNCEMENTS

(No announcements have been posted yet.)

# Calendar (8 sessions: 11am-12:30pm)

1	03/02	Introduction: So you want to be a linguist?
2	10/02	Translation processes and environments
3	17/02 <b>331</b>	Translation and CAT tools; localising a mobile app
		(Holidays)
4	02/03	Technical writing (start Assignment 1)
5	09/03	Introduction to interpreting (Assignment 1 due)
6	16/03	Interpreting: from theory to practice
7	06/04 <b>331</b>	Audiovisual translation: introduction to revoicing
8	18/03	Accessibility in practice (deadline to confirm research topic)
		(Assignment 2 deadline 04/05)

## UE 5 – Parcours de spécialisation – semestre 4

Parcours « Stratégie de marques et d'entreprises »

Communication et marketing  
(CM+TD)

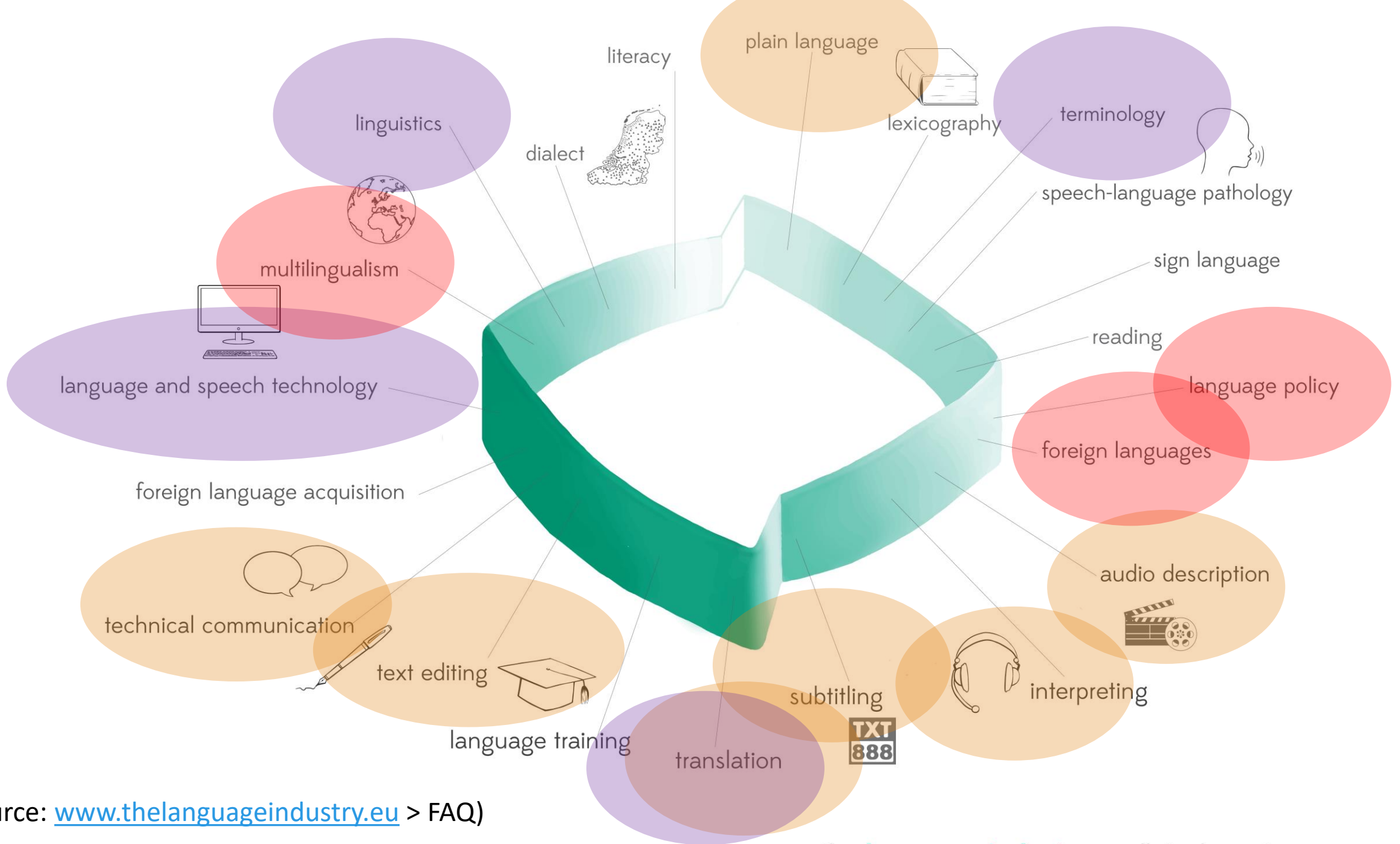
Comportement du consommateur  
(TD)

Parcours « Industries de la langue »

Métiers de la langue

*Translation studies* et linguistique appliquée

Langue et relations internationales



(Source: [www.thelanguageindustry.eu](http://www.thelanguageindustry.eu) > FAQ)

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.aal.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!

## Traduction

Les mots au kilo ?

Des normes au secours  
des acheteurs et des  
prestataires de services  
de traduction

Traduction  
faire les bons choix



Petit guide de l'acheteur de traductions



Guidelines on  
Gender-Neutral  
Language



# Case study: understanding machine translation

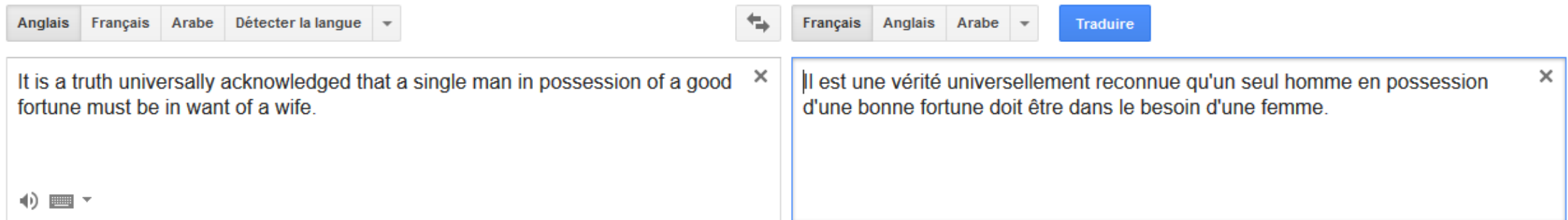
---

- 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)



The screenshot shows the Google Translate interface. On the left, the source language is set to 'Anglais' and the text is 'It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife.' On the right, the target language is set to 'Français' and the translated text is 'Il est une vérité universellement reconnue qu'un seul homme en possession d'une bonne fortune doit être dans le besoin d'une femme.' The interface includes a 'Traduire' button and a 'Détecter la langue' dropdown menu.

- 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)

The screenshot shows the Google Translate web interface. At the top left is the Google logo. To the right is a 'Connexion' button. Below the logo is a yellow banner with the text: 'Découvrez le navigateur Google Chrome doté d'une fonction de traduction automatique. [Télécharger Google Chrome](#) [Ignorer](#)'. The main heading is 'Traduction'. On the right side, there are 'G+' and '★' icons. The language selection area shows 'Anglais' selected for the source language and 'Français' selected for the target language. A 'Traduire' button is visible. The input text is 'All happy families are alike, but every unhappy family is unhappy in its own way.' The output text is 'Toutes les familles heureuses se ressemblent, mais chaque famille malheureuse est malheureuse à sa manière.' Below the input and output text are icons for voice input, keyboard input, and a 'Faux ?' button.

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)

Traduction



Anglais Français Arabe Détecter la langue



Français Anglais Chinois (simplifié)

Traduire

All happy families are alike, but every unhappy family is unhappy in its own way.

幸福的家庭都是相似的，但每个不幸的家庭各有各的不幸的方式。

Traduction



Anglais Français Chinois Détecter la langue



Français Anglais Chinois (simplifié)

Traduire

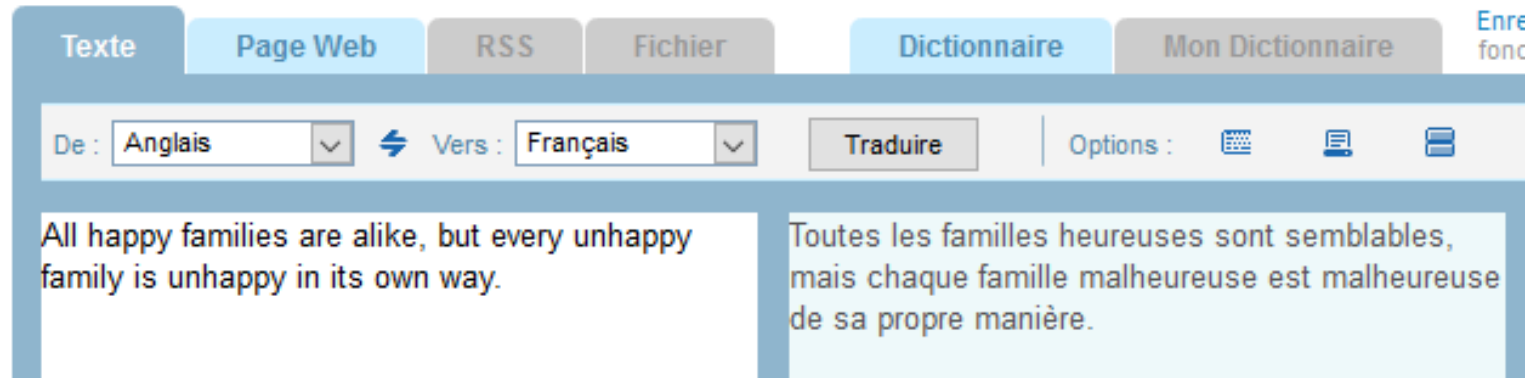
幸福的家庭都是相似的，但每个不幸的家庭各有各的不幸的方式。

Les familles heureuses se ressemblent tous, mais chaque famille malheureuse est malheureuse à sa manière.

Xìngfú de jiātíng dōu shì xiāngsì de, dàn měi gè bùxìng de jiātíng gè yǒu gè de bùxìng de fāngshì.

# 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)

Texte Page Web RSS Fichier Dictionnaire Mon Dictionnaire Enre fonc

De : Anglais Vers : Chinois Traduire Options :

All happy families are alike, but every unhappy family is unhappy it its own way. 所有愉快的家庭是象的，但是每个不快乐的家庭是不快乐的它它自己的方式。

象

De : Chinois Vers : Anglais Traduire Options :

所有愉快的家庭是象的,但是每个不快乐的家庭是不快乐的用它自己的方式。 All happy families are the elephants, but each not joyful family is not joyful uses its way.

- 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!

elephant



象



'eləfənt



Définitions de elephant

Nom

- ① a heavy plant-eating mammal with a prehensile trunk, long curved ivory tusks, and large ears, native to Africa and southern Asia. It is the largest living land animal.

"His notoriety first spread as the poacher of wild elephants for their precious tusks."

Nom

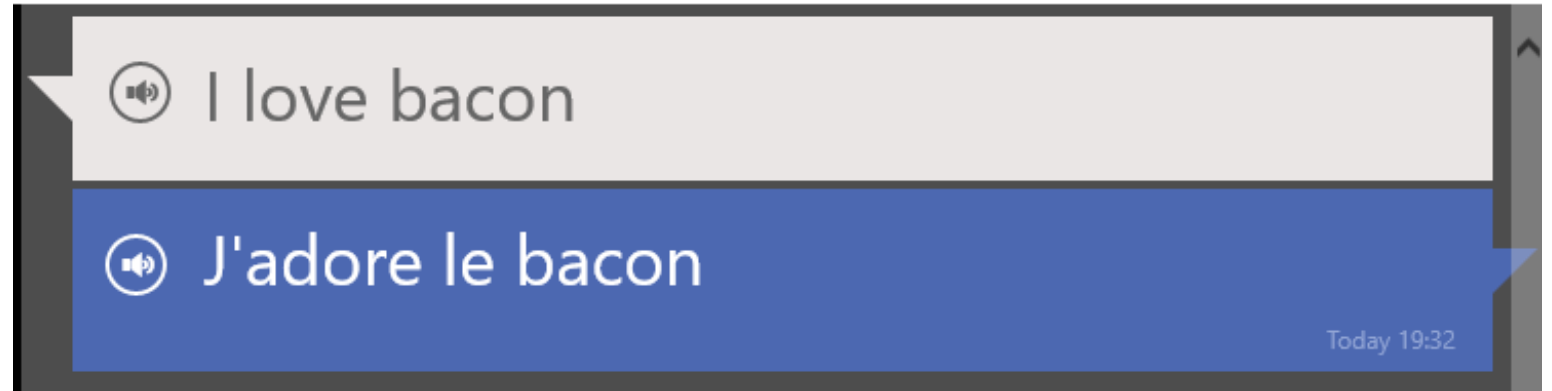
- 象 elephant, appearance, shape, jumbo
- 像 image, likeness, resemblance, elephant, appearance, similarity

Fréquence ⓘ



- ❑ **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):



# Machine translation + voice recognition (Microsoft Translator)

The screenshot shows the Microsoft Translator interface. The top section displays the English input: "It is a truth universally acknowledged a single man position of good fortune must be bored wife". Several words are highlighted in orange boxes, indicating errors: "a", "single", "man", "position", "of", "good", "fortune", "must", "be", "bored", and "wife". The bottom section displays the French output: "C'est qu'une vérité universellement reconnu qu'un seul homme de bonne fortune doit être bored femme". Several words are circled in orange, indicating errors: "qu'une", "reconnu", "un", "seul", "homme", "de", "bonne", "fortune", "doit", "être", "bored", and "femme". A small timestamp "Today 19:33" is visible in the bottom right corner of the interface.

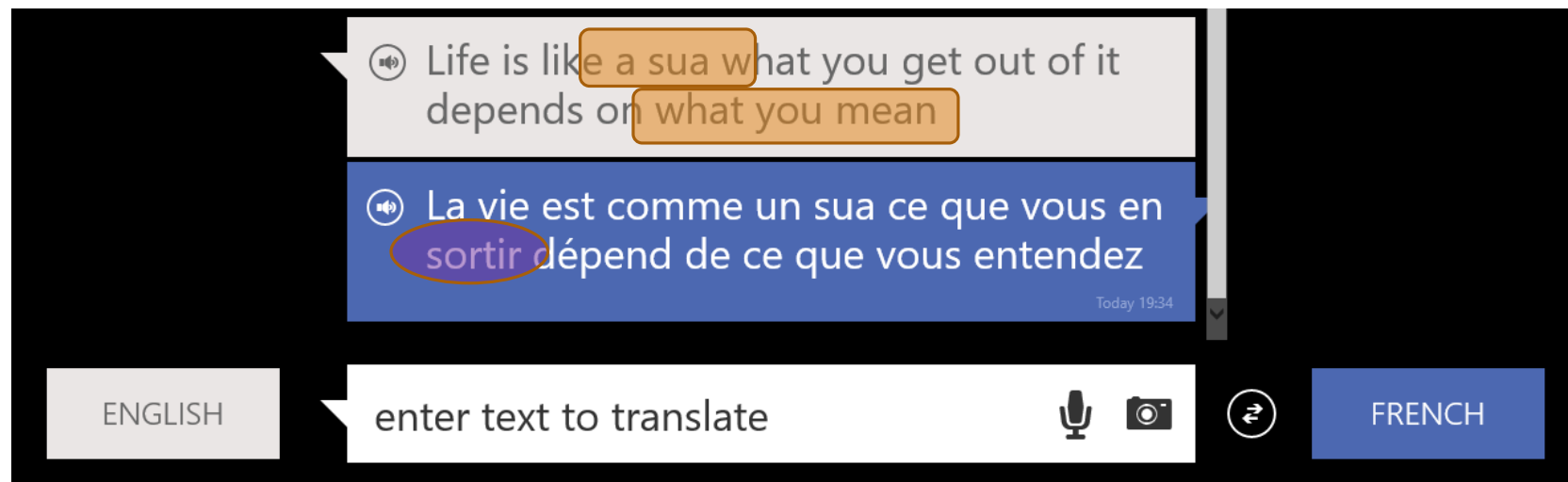
“It is a truth universally  
acknowledged that a single  
man in possession of a good  
fortune must be in want of a  
wife.”

(Jane Austen)

- ❑ Complex sentence with syntactical/grammatical complications:
  - ❑ impersonal subject: *It is*
  - ❑ unusual word order: *a truth <> universally acknowledged*
  - ❑ subordinate clause introduced by *that*
  - ❑ complex noun phrase as subject of SC:  
*a single man in possession of a good fortune*
  - ❑ complex verbal construction: *must be in want of*
- ❑ Speech recognition issues:
  - ❑ *that* unrecognised > /*thət*/ (vowel reduction)
  - ❑ *man in possession of* > \**man position of*  
reduction of ‘in’, vowel misrecognition, mis-analysis as compound noun \**man position*
  - ❑ *in want of a wife* > \**bored wife*  
mis-recognition of fixed expression *in want of*  
> *in* and *of* pronounced indistinctly when reduced  
> role of natural language **corpora** in recognition engine:  
*in want of a* not common in contemporary English  
*bored wife* an (apparently!) common web search term
- ❑ Machine translation issues:
  - ❑ *c’est qu’une* > syntax issue hinders sentence construction
  - ❑ *vérité ... reconnu* > N/Adj agreement issue
  - ❑ *un seul homme* > statistically more likely syntactic pattern for “a single <noun>”, but incorrect pragmatic choice here.
  - ❑ *bored femme* > ungrammatical ST = badly-formed TT!

## Machine translation + voice recognition (Microsoft Translator)

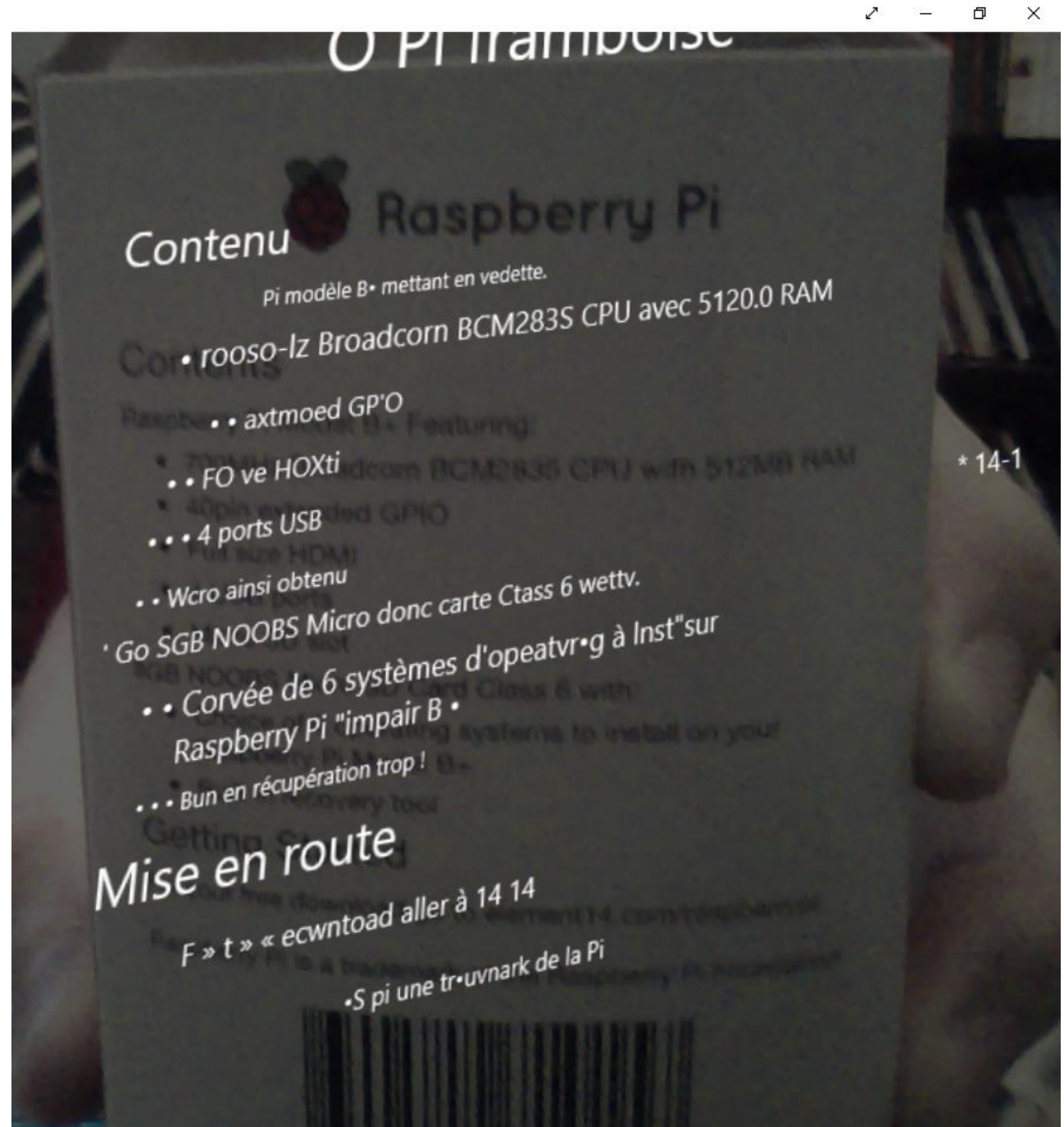
‘Life is like a sewer.  
What you get out of  
it depends on what  
you put in.’  
(Tom [Lehrer](#))



- ❑ Voice recognition issues:
  - ❑ Pause between sentences
  - ❑ Misrecognition of *sewer* as *\*sua*: pronounced /su:ə/ in non-rhotic Aus or UK English vs. /su:r/ or /su:ə/ 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.

# 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*



Translate text

Translate documents

English

Logout

# Machine translation and specialised domains – what's the problem here?

Là ci darem la mano  
Là mi dirai di sì  
Vedi, non è lontano  
Partiam, ben mio, da qui!  
Vorrei, et non vorrei  
Mi trema un poco il cor  
Felice, è ver, sarei,  
Ma può burlarmi ancor!



E-mail me my translation

174 / 2500



From

Italian



To

English

Domain

EU Formal Language



Translate text

Next to give me the hand  
Please tell me  
See, not far from.  
For the sake of completeness, I am right here!  
Let me, let me say  
I am a little bit



# 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](#) (online, create a free account) and/or [OmegaT](#) (free to download and install). We will be using these in class and for your first group assignment.**