
Python 3 : des fondamentaux aux concepts avancés du langage

Version 1.0

Thierry Parmentelat & Arnaud Legout

déc. 04, 2017

Sommaire:

| | | |
|-----------|---|-----------|
| 1 | Installer la distribution standard python | 3 |
| 1.1 | Complément - niveau basique | 3 |
| 2 | Digression - coexistence de python2 et python3 | 5 |
| 3 | Installation de base | 7 |
| 3.1 | Système | 7 |
| 4 | Anaconda | 9 |
| 5 | Un peu de lecture | 11 |
| 5.1 | Complément - niveau basique | 11 |
| 5.2 | Complément - niveau intermédiaire | 12 |
| 6 | « Notebooks » Jupyter comme support de cours | 13 |
| 6.1 | Notebooks | 13 |
| 7 | Modes d'exécution | 17 |
| 7.1 | Outils | 17 |
| 8 | La suite de Fibonacci | 21 |
| 8.1 | Complément - niveau basique | 21 |
| 9 | La suite de Fibonacci (suite) | 23 |
| 9.1 | Complément - niveau intermédiaire | 23 |
| 10 | La ligne <i>shebang</i> | 25 |
| 10.1 | Complément - niveau avancé | 25 |
| 11 | Dessiner un carré | 27 |
| 11.1 | Exercice - niveau intermédiaire | 27 |
| 11.2 | Exercice - niveau avancé | 28 |
| 12 | Noms de variables | 29 |
| 12.1 | Complément - niveau basique | 29 |
| 13 | Les mots-clés de python | 33 |
| 13.1 | En Python | 33 |
| 14 | Un peu de calcul sur les types | 35 |
| 14.1 | Complément - niveau basique | 35 |
| 14.2 | Complément - niveau avancé | 36 |

| | | |
|-----------|--|------------|
| 15 | Gestion de la mémoire | 37 |
| 15.1 | Complément - niveau basique | 37 |
| 15.2 | Complément - niveau intermédiaire | 37 |
| 16 | Typages statique et dynamique | 39 |
| 16.1 | Complément - niveau intermédiaire | 39 |
| 17 | Utiliser python comme une calculatrice | 43 |
| 17.1 | Dans l'interpréteur | 43 |
| 18 | Affectations & Opérations (à la +=) | 49 |
| 18.1 | Complément - niveau intermédiaire | 49 |
| 19 | Notions sur la précision des calculs flottants | 51 |
| 19.1 | Complément - niveau avancé | 51 |
| 20 | Opérations <i>bitwise</i> | 53 |
| 20.1 | Compléments - niveau avancé | 53 |
| 21 | Estimer le plus petit (grand) flottant | 57 |
| 21.1 | Exercice - niveau basique | 57 |
| 21.2 | Complément - niveau avancé | 59 |
| 22 | Caractères accentués | 61 |
| 22.1 | Complément - niveau basique | 61 |
| 22.2 | Complément - niveau intermédiaire | 62 |
| 23 | Les outils de base sur les strings | 67 |
| 23.1 | Complément - niveau intermédiaire | 67 |
| 24 | Formatage de chaînes de caractères | 71 |
| 24.1 | Complément - niveau basique | 71 |
| 24.2 | Complément - niveau intermédiaire | 74 |
| 24.3 | Complément - niveau avancé | 74 |
| 25 | Obtenir une réponse de l'utilisateur | 77 |
| 25.1 | Complément - niveau basique | 77 |
| 26 | Expressions régulières et le module <code>re</code> | 79 |
| 26.1 | Complément - niveau intermédiaire | 79 |
| 27 | Expressions régulières | 93 |
| 27.1 | Liens utiles | 93 |
| 27.2 | identificateurs python | 93 |
| 27.3 | lignes avec nom et prénom | 94 |
| 27.4 | numéros de téléphone | 94 |
| 28 | Les slices en python | 97 |
| 28.1 | Complément - niveau basique | 97 |
| 28.2 | Complément - niveau avancé | 100 |
| 29 | Méthodes spécifiques aux listes | 103 |
| 29.1 | Complément - niveau basique | 103 |
| 30 | Objets mutables et objets immuables | 109 |
| 30.1 | Complément - niveau basique | 109 |
| 31 | Tris de listes | 111 |
| 31.1 | Complément - niveau basique | 111 |

| | |
|--|------------|
| 32 Indentations en python | 115 |
| 32.1 Complément - niveau basique | 115 |
| 32.2 Complément - niveau intermédiaire | 116 |
| 32.3 Complément - niveau avancé | 117 |
| 33 Bonnes pratiques de présentation de code | 119 |
| 33.1 Complément - niveau basique | 119 |
| 33.2 Complément - niveau intermédiaire | 121 |
| 34 L'instruction pass | 123 |
| 34.1 Complément - niveau basique | 123 |
| 34.2 Complément - niveau intermédiaire | 124 |
| 35 Fonctions avec ou sans valeur de retour | 125 |
| 35.1 Complément - niveau basique | 125 |
| 36 Formatage | 129 |
| 36.1 Exercice - niveau basique | 129 |
| 37 Séquences | 131 |
| 37.1 Exercice - niveau basique | 131 |
| 37.2 Exercice - niveau intermédiaire | 132 |
| 38 Listes | 135 |
| 38.1 Exercice - niveau basique | 135 |
| 39 Instruction if et fonction def | 137 |
| 39.1 Exercice - niveau basique | 137 |
| 39.2 Exercice - niveau basique | 137 |
| 40 Compréhensions | 139 |
| 40.1 Exercice - niveau basique | 139 |
| 40.2 Récréation | 140 |
| 41 Compréhensions | 141 |
| 41.1 Exercice - niveau intermédiaire | 141 |
| 42 Les fichiers | 143 |
| 42.1 Complément - niveau basique | 143 |
| 42.2 Complément - niveau intermédiaire | 144 |
| 42.3 Complément - niveau avancé | 145 |
| 43 Fichiers et utilitaires | 149 |
| 43.1 Complément - niveau basique | 149 |
| 43.2 Complément - niveau avancé | 152 |
| 44 Formats de fichiers : JSON et autres | 155 |
| 44.1 Compléments - niveau basique | 155 |
| 44.2 Compléments - niveau intermédiaire | 156 |
| 45 Fichiers systèmes | 159 |
| 45.1 Complément - niveau avancé | 159 |
| 46 La construction de tuples | 163 |
| 46.1 Complément - niveau intermédiaire | 163 |
| 47 Sequence unpacking | 167 |
| 47.1 Complément - niveau basique | 167 |
| 47.2 Complément - niveau intermédiaire | 169 |
| 47.3 Pour en savoir plus | 171 |

| | |
|---|------------|
| 48 Plusieurs variables dans une boucle <code>for</code> | 173 |
| 48.1 Complément - niveau basique | 173 |
| 48.2 Complément - niveau intermédiaire | 173 |
| 49 Fichiers | 177 |
| 49.1 Exercice - niveau basique | 177 |
| 50 <i>Sequence unpacking</i> | 179 |
| 50.1 Exercice - niveau basique | 179 |
| 51 Dictionnaires | 181 |
| 51.1 Complément - niveau basique | 181 |
| 51.2 Complément - niveau intermédiaire | 183 |
| 51.3 Complément - niveau avancé | 186 |
| 52 Clés immuables | 189 |
| 52.1 Complément - niveau intermédiaire | 189 |
| 53 Gérer des enregistrements | 193 |
| 53.1 Complément - niveau intermédiaire | 193 |
| 53.2 Complément - niveau avancé | 195 |
| 54 Dictionnaires et listes | 197 |
| 54.1 Exercice - niveau basique | 197 |
| 55 Fusionner des données | 199 |
| 55.1 Exercices | 199 |
| 56 Ensembles | 205 |
| 56.1 Complément - niveau basique | 205 |
| 57 Ensembles | 211 |
| 57.1 Exercice - niveau basique | 211 |
| 57.2 Deuxième partie - niveau basique | 212 |
| 58 Exercice sur les ensembles | 215 |
| 58.1 Exercice - niveau intermédiaire | 215 |
| 59 <code>try .. else .. finally</code> | 219 |
| 59.1 Complément - niveau intermédiaire | 219 |
| 60 L'opérateur <code>is</code> | 223 |
| 60.1 Complément - niveau basique | 223 |
| 60.2 Complément - niveau intermédiaire | 224 |
| 61 Listes infinies & références circulaires | 227 |
| 61.1 Complément - niveau intermédiaire | 227 |
| 62 Les différentes copies | 231 |
| 62.1 Complément - niveau basique | 231 |
| 62.2 Complément - niveau intermédiaire | 233 |
| 63 L'instruction <code>del</code> | 237 |
| 63.1 Complément - niveau basique | 237 |
| 64 Affectation simultanée | 241 |
| 64.1 Complément - niveau basique | 241 |
| 65 Les instructions <code>+=</code> et autres revisitées | 243 |
| 65.1 Complément - niveau intermédiaire | 243 |

| | |
|--|------------|
| 66 Classe | 247 |
| 66.1 Exercice - niveau basique | 247 |
| 67 Passage d'arguments par référence | 249 |
| 67.1 Complément - niveau intermédiaire | 249 |
| 68 Rappels sur <i>docstring</i> | 253 |
| 68.1 Complément - niveau basique | 253 |
| 69 <i>isinstance</i> | 257 |
| 69.1 Complément - niveau basique | 257 |
| 69.2 Complément - niveau intermédiaire | 258 |
| 70 <i>Type hints</i> | 261 |
| 70.1 Complément - niveau intermédiaire | 261 |
| 70.2 Complément - niveau avancé | 264 |
| 71 Conditions & Expressions Booléennes | 267 |
| 71.1 Complément : niveau basique | 267 |
| 72 Évaluation des tests | 271 |
| 72.1 Complément - niveau basique | 271 |
| 72.2 Complément - niveau intermédiaire | 272 |
| 73 Une forme alternative du <i>if</i> | 275 |
| 73.1 Complément - niveau basique | 275 |
| 73.2 Complément - niveau intermédiaire | 276 |
| 74 Récapitulatif sur les conditions dans un <i>if</i> | 277 |
| 74.1 Complément - niveau basique | 277 |
| 74.2 Complément - niveau intermédiaire | 280 |
| 75 L'instruction <i>if</i> | 283 |
| 75.1 Exercice - niveau basique | 283 |
| 75.2 Exercice - niveau basique | 283 |
| 76 Expression conditionnelle | 285 |
| 76.1 Exercice - niveau basique | 285 |
| 77 La boucle <i>while .. else</i> | 287 |
| 77.1 Complément - niveau basique | 287 |
| 77.2 Complément - niveau intermédiaire | 287 |
| 78 Calculer le PGCD | 289 |
| 78.1 Exercice - niveau basique | 289 |
| 79 Exercice | 291 |
| 79.1 Niveau basique | 291 |
| 80 Le scope <i>builtin</i> | 293 |
| 80.1 Complément - niveau avancé | 293 |
| 81 Visibilité des variables de boucle | 295 |
| 81.1 Complément - niveau basique | 295 |
| 82 L'exception <i>UnboundLocalError</i> | 299 |
| 82.1 Complément - niveau intermédiaire | 299 |
| 83 Les fonctions <i>globals</i> et <i>locals</i> | 303 |
| 83.1 Complément - niveau intermédiaire | 303 |

| | | |
|-----------|---|------------|
| 84 | Passage d'arguments | 307 |
| 84.1 | Complément - niveau intermédiaire | 307 |
| 84.2 | Complément - niveau avancé | 309 |
| 85 | Un piège courant | 313 |
| 85.1 | Complément - niveau basique | 313 |
| 85.2 | Complément - niveau intermédiaire | 313 |
| 86 | Arguments <i>keyword-only</i> | 315 |
| 86.1 | Complément - niveau intermédiaire | 315 |
| 87 | Passage d'arguments | 317 |
| 87.1 | Exercice - niveau basique | 317 |
| 87.2 | Exercice - niveau intermédiaire | 317 |

Python 3 : des fondamentaux aux concepts avancés du langage

Licence CC BY-NC-ND Thierry Parmentelat & Arnaud Legout



Fun MOOC : <https://www.fun-mooc.fr>

Installer la distribution standard python

1.1 Complément - niveau basique

Ce complément a pour but de vous donner quelques guides pour l'installation de la distribution standard python 3.

Notez bien qu'il ne s'agit ici que d'indications, il existe de nombreuses façons de procéder.

En cas de souci, commencez par chercher par vous-même sur google ou autre une solution à votre problème ; pensez également à utiliser le forum du cours.

Le point important est de **bien vérifier le numéro de version** de votre installation qui doit être **au moins 3.6**

Digression - coexistence de python2 et python3

Avant l'arrivée de la version 3 de python, les choses étaient simples, on exécutait un programme python avec une commande `python`. Depuis 2014-2015, maintenant que les deux versions de python coexistent, il est nécessaire d'adopter une convention qui permette d'installer les deux langages sous des noms qui sont non-ambigus.

C'est pourquoi actuellement, on trouve **le plus souvent** la convention suivante sous Linux et Mac OS X :

- `python3` est pour exécuter les programmes en python-3; du coup on trouve alors également les commandes comme `idle3` pour lancer IDLE, et par exemple `pip3` pour le gestionnaire de paquets (voir ci-dessous).
- `python2` est pour exécuter les programmes en python-2, avec typiquement `idle2` et `pip2`;
- enfin selon les systèmes, la commande `python` tout court est un alias pour `python2` ou `python3`. De plus en plus souvent, par défaut `python` désigne `python3`.

à titre d'illustration, voici ce que j'obtiens sur mon mac :

```
$ python3 -V
Python 3.6.2
$ python2 -V
Python 2.7.13
$ python -V
Python 3.6.2
```

Sous Windows, vous avez un lanceur qui s'appelle `py`. Par défaut, il lance la version de python la plus récente installée, mais vous pouvez spécifier une version spécifique de la manière suivante :

```
C:\> py -2.7
```

pour lancer, par exemple, python en version 2.7. Vous trouverez toute la documentation nécessaire pour Windows sur cette page (en anglais) : <https://docs.python.org/3/using/windows.html>

Pour éviter d'éventuelles confusions, nous précisons toujours `python3` dans le cours.

Installation de base

3.1 Système

3.1.1 Vous utilisez Windows

La méthode recommandée sur Windows est de partir de la page <https://www.python.org/download> où vous trouverez un programme d'installation qui contient tout ce dont vous aurez besoin pour suivre le cours.

Pour vérifier que vous êtes prêts, il vous faut lancer IDLE (quelque part dans le menu Démarrer) et vérifier le numéro de version.

3.1.2 Vous utilisez MacOS

Ici encore, la méthode recommandée est de partir de la page <https://www.python.org/download> et d'utiliser le programme d'installation.

Sachez aussi, si vous utilisez déjà MacPorts (<https://www.macports.org>), que vous pouvez également utiliser cet outil pour installer, par exemple python 3.6, avec la commande

```
$ sudo port install python36
```

3.1.3 Vous utilisez Linux

Dans ce cas il y est très probable que python-3.x est déjà disponible sur votre machine. Pour vous en assurer, essayez de lancer la commande `python3` dans un terminal.

Redhat / Fedora

Voici par exemple ce qu'on obtient depuis un terminal sur une machine installée en Fedora-20

```
$ python3
Python 3.6.2 (default, Jul 20 2017, 12:30:02)
[GCC 6.3.1 20161221 (Red Hat 6.3.1-1)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
```

Vérifiez bien le numéro de version qui doit être en 3.x. Si vous obtenez un message du style `python3: command not found` utilisez `dnf` (anciennement connu sous le nom de `yum`) pour installer le rpm `python3` comme ceci

```
$ sudo dnf install python3
```

S'agissant de `idle`, l'éditeur que nous utilisons dans le cours (optionnel si vous êtes familier avec un éditeur de texte), vérifiez sa présence comme ceci

```
$ type idle3
idle is hashed (/usr/bin/idle3)
```

Ici encore, si la commande n'est pas disponible vous pouvez l'installer avec

```
$ sudo yum install python3-tools
```

Debian / Ubuntu

Ici encore, `python-2.7` est sans doute déjà disponible. Procédez comme ci-dessus, voici un exemple recueilli dans un terminal sur une machine installée en Ubuntu-14.04/trusty

```
$ python3
Python 3.6.2 (default, Jul 20 2017, 12:30:02)
[GCC 6.3.1 20161221 (Red Hat 6.3.1-1)] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()
```

Pour installer python

```
$ sudo apt-get install python3
```

Pour installer idle

```
$ sudo apt-get install idle3
```

3.1.4 Installation de librairies complémentaires

Il existe un outil très pratique pour installer les librairies python, il s'appelle `pip3`, qui est documenté ici <https://pypi.python.org/pypi/pip>

Sachez aussi, si par ailleurs vous utilisez un gestionnaire de package comme `rpm` sur RedHat, `apt-get` sur debian, ou `port` sur MacOS, que de nombreux packages sont également disponibles au travers de ces outils.

CHAPITRE 4

Anaconda

Sachez qu'il existe beaucoup de distributions alternatives qui incluent python ; parmi elles, la plus populaire est sans aucun doute [Anaconda](#), qui contient un grand nombre de bibliothèques de calcul scientifique, et également d'ailleurs jupyter pour travailler nativement sur des notebooks au format `.ipynb`.

Anaconda vient avec son propre gestionnaires de paquets pour l'installation de librairies supplémentaires qui s'appelle `conda`.

5.1 Complément - niveau basique

5.1.1 Le zen de python

Vous pouvez lire le « zen de python », qui résume la philosophie du langage, en important le module `this` avec ce code : (pour exécuter ce code, cliquer dans la cellule de code, et faites au clavier « Majuscule/Entrée » ou « Shift/Enter »)

```
# le zen de python
import this
```

The Zen of Python, by Tim Peters

```
Beautiful is better than ugly.
Explicit is better than implicit.
Simple is better than complex.
Complex is better than complicated.
Flat is better than nested.
Sparse is better than dense.
Readability counts.
Special cases aren't special enough to break the rules.
Although practicality beats purity.
Errors should never pass silently.
Unless explicitly silenced.
In the face of ambiguity, refuse the temptation to guess.
There should be one-- and preferably only one --obvious way to do it.
Although that way may not be obvious at first unless you're Dutch.
Now is better than never.
Although never is often better than right now.
If the implementation is hard to explain, it's a bad idea.
If the implementation is easy to explain, it may be a good idea.
Namespaces are one honking great idea -- let's do more of those!
```

5.1.2 Documentation

- On peut commencer par citer l'article de wikipedia sur python en français
- La page sur le langage en français
- La documentation originale de python-3 - donc, en anglais - est un très bon point d'entrée lorsqu'on cherche un sujet particulier, mais (beaucoup) trop abondante pour être lue d'un seul trait. Pour chercher de la documentation sur un module particulier, le plus simple est encore d'utiliser google - ou votre moteur de recherche favori - qui vous redirigera dans la grande majorité des cas vers la page qui va bien dans, précisément, la documentation python.

À titre d'exercice, cherchez la documentation du module `pathlib` en cherchant sur google « *'python module pathlib'* » >> <https://www.google.fr/search?q=python+module+pathlib> >__.

5.1.3 Historique et survol

- La FAQ officielle de Python (en anglais) sur les choix de conception et l'historique du langage
- L'article de wikipedia (en anglais) sur l'historique du langage
- Sur wikipédia, un article (en anglais) sur la syntaxe et la sémantique de python

5.1.4 Un peu de folklore

- Le talk de Guido van Rossum à PyCon 2016
- Sur youtube, le sketch des monty python d'où provient les termes spam, eggs et autres beans qu'on utilise traditionnellement dans les exemples en python plutôt que `foo` et `bar`
- L'article wikipedia correspondant, qui cite le langage python

5.2 Complément - niveau intermédiaire

5.2.1 Licence

- La licence d'utilisation est disponible ici
- La page de la Python Software Foundation, qui est une entité légale similaire à nos associations de 1901, à but non lucratif; elle possède les droits sur le langage

5.2.2 Le processus de développement

- Comment les choix d'évolution sont proposés et discutés, au travers des PEP (Python Enhancement Proposal)
- [http://en.wikipedia.org/wiki/Python_\(programming_language\)#Development](http://en.wikipedia.org/wiki/Python_(programming_language)#Development)
- Le premier PEP décrit en détail le cycle de vie des PEPs
- <http://legacy.python.org/dev/peps/pep-0001/>
- Le PEP 8, qui préconise un style de présentation (*style guide*)
- <http://legacy.python.org/dev/peps/pep-0008/>
- L'index de tous les PEPs
- <http://legacy.python.org/dev/peps/>

« Notebooks » Jupyter comme support de cours

6.1 Notebooks

Pour illustrer les vidéos du MOOC, nous avons choisi d'utiliser Jupyter pour vous rédiger les documents « mixtes » contenant du texte et du code python, qu'on appelle des « notebooks », et dont le présent document est un exemple.

Nous allons dans la suite utiliser du code Python, pourtant nous n'avons pas encore abordé le langage. Pas d'inquiétude, ce code est uniquement destiné à valider le fonctionnement des notebooks, et nous n'utilisons que des choses très simples.

6.1.1 Avantages des notebooks

Comme vous le voyez, ce support permet un format plus lisible que des commentaires dans un fichier de code.

Nous attirons votre attention sur le fait que **les fragments de code peuvent être évalués et modifiés**. Ainsi vous pouvez facilement essayer des variantes autour du notebook original.

Notez bien également que le code python est interprété **sur une machine distante**, ce qui vous permet de faire vos premiers pas avant même d'avoir procédé à l'installation de python sur votre propre ordinateur.

6.1.2 Comment utiliser les notebooks

En haut du notebook, vous avez une barre, contenant :

- un titre pour le notebook, avec un numéro de version,
- une barre de menus avec les entrées `File`, `Edit`, `View`, `Insert` ... ,
- et une barre de boutons qui sont des raccourcis vers certains menus fréquemment utilisés. Si vous laissez votre souris au dessus d'un bouton, un petit texte apparaît, indiquant à quelle fonction correspond ce bouton.

Nous avons vu dans la vidéo qu'un notebook est constitué d'une suite de cellules, soit textuelles, soit contenant du code. Les cellules de code sont facilement reconnaissables, elles sont précédées de `In [] :`. La cellule qui suit celle que vous êtes en train de lire est une cellule de code.

Pour commencer, sélectionnez cette cellule de code avec votre souris, et appuyez dans la barre de boutons sur celui en forme de flèche triangulaire vers la droite (Play)

```
20 * 30
```

```
600
```

Comme vous le voyez, la cellule est « exécutée » (on dira plus volontiers évaluée), et on passe à la cellule suivante.

Alternativement vous pouvez simplement taper au clavier **Shift+Enter** ou selon les claviers **Maj-Entrée**, pour obtenir le même effet. D'une manière générale, il est important d'apprendre et d'utiliser les raccourcis clavier, cela vous fera gagner beaucoup de temps par la suite.

La façon habituelle d'*exécuter* l'ensemble du notebook consiste à partir de la première cellule, et à taper **Shift+Enter** jusqu'au bout du notebook.

Lorsqu'une cellule de code a été évaluée, Jupyter ajoute sous la cellule In une cellule Out qui donne le résultat du fragment python, soit ci-dessus 600.

Jupyter ajoute également un nombre entre les crochets pour afficher, par exemple ci-dessus, In [1] : . Ce nombre vous permet de retrouver l'ordre dans lequel les cellules ont été évaluées.

Vous pouvez naturellement modifier ces cellules de code pour faire des essais ; ainsi vous pouvez vous servir du modèle ci-dessous pour calculer la racine carrée de 3, ou essayer la fonction sur un nombre négatif et voir comment est signalée l'erreur.

```
# math.sqrt (pour square root) calcule la racine carrée
import math
math.sqrt(2)
```

```
1.4142135623730951
```

On peut également évaluer tout le notebook en une seule fois en utilisant le menu *Cell -> Run All*

6.1.3 Attention à bien évaluer les cellules dans l'ordre

Il est important que les cellules de code soient évaluées dans le bon ordre. Si vous ne respectez pas l'ordre dans lequel les cellules de code sont présentées, le résultat peut être inattendu.

En fait, évaluer un programme sous forme de notebook revient à le découper en petits fragments, et si on exécute ces fragments dans le désordre, on obtient naturellement un programme différent.

On le voit sur cet exemple

```
message = "Il faut faire attention à l'ordre dans lequel on évalue les notebooks"
```

```
print(message)
```

```
Il faut faire attention à l'ordre dans lequel on évalue les notebooks
```

Si un peu plus loin dans le notebook on fait par exemple

```
# ceci a pour effet d'effacer la variable 'message'
del message
```

qui rend le symbole « message » indéfini, alors bien sûr on ne peut plus évaluer la cellule qui fait `print` puisque la variable `message` n'est plus connue de l'interpréteur.

6.1.4 Réinitialiser l'interpréteur

Si vous faites trop de modifications, ou perdez le fil de ce que vous avez évalué, il peut être utile de redémarrer votre interpréteur. Le menu *Kernel -> Restart* vous permet de faire cela, un peu à la manière de IDLE qui repart d'un interpréteur vierge lorsque vous utilisez la fonction F5.

Le menu *Kernel* → *Interrupt* peut être quant à lui utilisé si votre fragment prend trop longtemps à s'exécuter (par exemple vous avez écrit une boucle dont la logique est cassée et qui ne termine pas).

6.1.5 Vous travaillez sur une copie

Un des avantages principaux des notebooks est de vous permettre de modifier le code que nous avons écrit, et de voir par vous mêmes comment se comporte le code modifié.

Pour cette raison, chaque élève dispose de sa **propre copie** de chaque notebook, vous pouvez bien sûr apporter toutes les modifications que vous souhaitez à vos notebooks sans affecter les autres étudiants.

6.1.6 Revenir à la version du cours

Vous pouvez toujours revenir à la version « du cours » grâce au menu *File* → *Reset to original*

Attention, avec cette fonction vous restaurez **tout le notebook** et donc **vous perdez vos modifications sur ce notebook**.

6.1.7 Télécharger au format python

Vous pouvez télécharger un notebook au format python sur votre ordinateur grâce au menu *File* → *Download as* → *python*

Les cellules de texte sont préservées dans le résultat sous forme de commentaires python.

6.1.8 Partager un notebook en lecture seule

Enfin, avec le menu *File* → *Share static version*, vous pouvez publier une version en lecture seule de votre notebook ; vous obtenez une URL que vous pouvez publier par exemple pour demander de l'aide sur le forum. Ainsi, les autres étudiants peuvent accéder en lecture seule à votre code.

Notez que lorsque vous utilisez cette fonction plusieurs fois, c'est toujours la dernière version publiée que verront vos camarades, l'URL utilisée reste toujours la même pour un étudiant et un notebook donné.

6.1.9 Ajouter des cellules

Vous pouvez ajouter une cellule n'importe où dans le document avec le bouton + de la barre de boutons.

Aussi, lorsque vous arrivez à la fin du document, une nouvelle cellule est créée chaque fois que vous évaluez la dernière cellule ; de cette façon vous disposez d'un brouillon pour vos propres essais.

À vous de jouer.

CHAPITRE 7

Modes d'exécution

Nous avons donc à notre disposition plusieurs façons d'exécuter un programme python. Nous allons les étudier plus en détail :

| Quoi | Avec quel outil |
|-----------------|---|
| fichier complet | <code>python3 <fichier>.py</code> |
| ligne à ligne | <code>python3</code> en mode interactif |
| | ou sous <code>ipython3</code> |
| | ou avec <code>IDLE</code> |
| par fragments | dans un notebook |

Pour cela nous allons voir le comportement d'un tout petit programme python lorsqu'on l'exécute sous ces différents environnements.

On veut surtout expliquer une petite différence quant au niveau de détail de ce qui se trouve imprimé.

Essentiellement, lorsqu'on utilise l'interpréteur en mode interactif - ou sous `IDLE` - à chaque fois que l'on tape une ligne, le résultat est **calculé** (on dit aussi **évalué**) puis **imprimé**.

Par contre, lorsqu'on écrit tout un programme, on ne peut plus imprimer le résultat de toutes les lignes, cela produirait un flot d'impression beaucoup trop important. Par conséquent, si vous ne déclenchez pas une impression avec, par exemple, la fonction `print`, rien ne s'affichera.

Enfin, en ce qui concerne le notebook, le comportement est un peu hybride entre les deux, en ce sens que seul le **dernier résultat** de la cellule est imprimé.

7.1 Outils

7.1.1 L'interpréteur python interactif

Le programme choisi est très simple, c'est le suivant

```
10 * 10
20 * 20
30 * 30
```

Voici comment se comporte l'interpréteur interactif quand on lui soumet ces instructions

```
$ python3
Python 3.5.1 (v3.5.1:37a07cee5969, Dec 5 2015, 21:12:44)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> 10 * 10
100
>>> 20 * 20
400
>>> 30 * 30
900
>>> exit()
$
```

Notez que pour terminer la session, il nous faut « sortir » de l'interpréteur en tapant `exit()`

On peut aussi taper `Control-D` sous linux ou MacOS.

Comme on le voit ici, l'interpréteur imprime **le résultat de chaque ligne**. On voit bien apparaître toutes les valeurs calculées, 100, 400, puis enfin 900.

7.1.2 Sous forme de programme constitué

Voyons à présent ce que donne cette même séquence de calculs dans un programme complet. Pour cela, il nous faut tout d'abord fabriquer un fichier avec un suffixe en `.py`, en utilisant par exemple un éditeur de fichier. Le résultat doit ressembler à ceci :

```
$ cat foo.py
10 * 10
20 * 20
30 * 30
$
```

Exécutons à présent ce programme :

```
$ python foo.py
$
```

On constate donc que ce programme **ne fait rien** ! En tous cas, selon toute apparence.

En réalité, les 3 valeurs 100, 400 et 900 sont bien calculées, mais comme aucune instruction `print` n'est présente, rien n'est imprimé et le programme se termine sans signe apparent d'avoir réellement fonctionné.

Ce comportement peut paraître un peu déroutant au début, mais comme nous l'avons mentionné c'est tout à fait délibéré. Un programme fonctionnel faisant facilement plusieurs milliers de lignes, voire beaucoup plus, il ne serait pas du tout réaliste que chaque ligne produise une impression, comme c'est le cas en mode interactif.

7.1.3 Dans un notebook

Voici à présent le même programme dans un notebook

```
10 * 10
20 * 20
30 * 30
```

```
900
```

Lorsqu'on exécute cette cellule (rappel : sélectionner la cellule, et utiliser le bouton en forme de flèche vers la droite, ou entrer « **Shift+Enter** » au clavier), on obtient une seule valeur dans la rubrique "Out", 900, qui correspond **au résultat de la dernière ligne**.

7.1.4 Utiliser `print`

Ainsi, pour afficher un résultat intermédiaire, on utilise l'instruction `print`. Nous verrons cette instruction en détail dans les semaines qui viennent, mais en guise d'introduction disons seulement que c'est une fonction comme les autres en python-3.

```
a = 10
b = 20

print(a, b)
```

```
10 20
```

On peut naturellement mélanger des objets de plusieurs types, et donc mélanger des strings et des nombres pour obtenir un résultat un peu plus lisible. En effet, lorsque le programme devient gros, il est important de savoir à quoi correspond une ligne dans le flot de toutes les impressions. Aussi on préférera quelque chose comme :

```
print("a =", a, "et b =", b)
```

```
a = 10 et b = 20
```

```
# ou encore, équivalente mais avec un f-string
print(f"a = {a} et b = {b}")
```

```
a = 10 et b = 20
```

Une pratique courante consiste d'ailleurs à utiliser les commentaires pour laisser dans le code les instructions `print` qui correspondent à du debug (c'est-à-dire qui ont pu être utiles lors de la mise au point et qu'on veut pouvoir réactiver rapidement).

7.1.5 Utiliser `print` pour « sous-titrer » une affectation

Remarquons enfin que l'affectation à une variable ne retourne aucun résultat.

C'est à dire, en pratique, que si on écrit

```
a = 100
```

même une fois l'expression évaluée par l'interpréteur, aucune ligne `Out []` n'est ajoutée.

C'est pourquoi, il nous arrivera parfois d'écrire alors plutôt, et notamment lorsque l'expression est complexe et pour rendre explicite la valeur qui vient d'être affectée

```
a = 100 ; print(a)
```

```
100
```

Notez bien que cette technique est uniquement pédagogique et n'a absolument aucun autre intérêt dans la pratique, il n'est **pas recommandé** de l'utiliser en dehors de ce contexte.

La suite de Fibonacci

8.1 Complément - niveau basique

Voici un premier exemple de code qui tourne.

Nous allons commencer par le faire tourner dans ce notebook. Nous verrons en fin de séance comment le faire fonctionner localement sur votre ordinateur.

Le but de ce programme est de calculer la [suite de Fibonacci](#), qui est définie comme ceci :

$$\begin{aligned} u_0 &= 1 \\ u_1 &= 1 \\ \forall n \geq 2, u_n &= u_{n-1} + u_{n-2} \end{aligned}$$

Ce qui donne pour les premières valeurs

| n | fibonacci(n) |
|---|--------------|
| 0 | 1 |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |
| 4 | 5 |
| 5 | 8 |
| 6 | 13 |

On commence par définir la fonction `fibonacci` comme suit. Naturellement vous n'avez pas encore tout le bagage pour lire ce code, ne vous inquiétez pas, nous allons vous expliquer tout ça dans les prochaines semaines. Le but est uniquement de vous montrer un fonctionnement de l'interpréteur Python et de IDLE.

```
def fibonacci(n):  
    "retourne le nombre de fibonacci pour l'entier n"  
    # pour les petites valeurs de n il n'y a rien à calculer  
    if n <= 1:  
        return 1  
    # sinon on initialise f1 pour n-1 et f2 pour n-2  
    f2, f1 = 1, 1  
    # et on itère n-1 fois pour additionner  
    for i in range(2, n + 1):
```

```
f2, f1 = f1, f1 + f2
#     print(i, f2, f1)
# le résultat est dans f1
return f1
```

Pour en faire un programme utilisable on va demander à l'utilisateur de rentrer un nombre ; il faut le convertir en entier car `input` renvoie une chaîne de caractères

```
entier = int(input("Entrer un entier "))
```

```
Entrer un entier 50
```

On imprime le résultat

```
print(f"fibonacci({entier}) = {fibonacci(entier)}")
```

```
fibonacci(50) = 20365011074
```

8.1.1 Exercice

Vous pouvez donc à présent :

- exécuter le code dans ce notebook
- télécharger ce code sur votre disque comme un fichier `fibonacci_prompt.py`
 - utilisez pour cela le menu « *File -> Download as -> python* »
 - et **renommez le fichier obtenu** au besoin
- l'exécuter sous IDLE
- le modifier, par exemple pour afficher les résultats intermédiaires
 - on a laissé exprès des fonctions `print` en commentaires que vous pouvez réactiver simplement
- l'exécuter avec l'interpréteur python comme ceci

```
$ python3 fibonacci_prompt.py
```

Ce code est volontairement simple et peu robuste pour ne pas l'alourdir. Par exemple, ce programme se comporte mal si vous entrez un entier négatif.

Nous allons voir tout de suite une version légèrement différente qui va vous permettre de donner la valeur d'entrée sur la ligne de commande.

La suite de Fibonacci (suite)

9.1 Complément - niveau intermédiaire

Nous reprenons le cas de la fonction `fibonacci` que nous avons vue déjà, mais cette fois nous voulons que l'utilisateur puisse indiquer l'entier en entrée de l'algorithme, non plus en répondant à une question, mais sur la ligne de commande, c'est-à-dire en tapant

```
$ python3 fibonacci.py 12
```

Avertissement :

Attention, cette version-ci **ne fonctionne pas dans ce notebook**, justement car on n'a pas de moyen dans un notebook d'invoquer un programme en lui passant des arguments de cette façon. Ce notebook est rédigé pour vous permettre de vous entraîner avec la fonction de téléchargement au format python, qu'on a vue dans la vidéo, et de faire tourner ce programme sur votre propre ordinateur.

9.1.1 Le module `argparse`

Cette fois nous importons le module `argparse`, c'est lui qui va nous permettre d'interpréter les arguments passés à la ligne de commande.

```
from argparse import ArgumentParser
```

Puis nous répétons la fonction `fibonacci`

```
def fibonacci(n):
    "retourne le nombre de fibonacci pour l'entier n"
    # pour les petites valeurs de n il n'y a rien à calculer
    if n <= 1:
        return 1
    # sinon on initialise f1 pour n-1 et f2 pour n-2
    f2, f1 = 1, 1
    # et on itère n-1 fois pour additionner
    for i in range(2, n + 1):
        f2, f1 = f1, f1 + f2
    #     print(i, f2, f1)
    # le résultat est dans f1
    return f1
```

Remarque :

Certains d'entre vous auront évidemment remarqué qu'on aurait pu éviter de copier-coller la fonction `fibonacci` comme cela; c'est à ça que servent les modules, mais nous n'en sommes pas là.

9.1.2 Un objet parser

À présent, nous utilisons le module `argparse`, pour lui dire qu'on attend exactement un argument sur la ligne de commande, et qui doit être un entier. Ici encore ne vous inquiétez pas si vous ne comprenez pas tout le code, l'objectif est de vous donner un morceau de code utilisable tout de suite pour jouer avec votre interpréteur python.

```
# à nouveau : ceci n'est pas conçu pour être exécuté dans le notebook !
parser = ArgumentParser()
parser.add_argument(dest="entier", type=int,
                    help="entier d'entrée")
input_args = parser.parse_args()
entier = input_args.entier
```

Nous pouvons à présent afficher le résultat

```
print(f"fibonacci({entier}) = {fibonacci(entier)}")
```

Vous pouvez donc à présent

- télécharger ce code sur votre disque comme un fichier `fibonacci.py` en utilisant le menu « *File -> Download as -> python* »
- l'exécuter avec simplement python comme ceci

```
$ python fibonacci.py 56
```


CHAPITRE 10

La ligne *shebang*

```
#!/usr/bin/env python3
```

10.1 Complément - niveau avancé

Ce complément est uniquement valable pour MacOS et linux.

10.1.1 Le besoin

Nous avons vu dans la vidéo que pour lancer un programme python on fait essentiellement depuis le terminal

```
$ python3 mon_module.py
```

Lorsqu'il s'agit d'un programme qu'on utilise fréquemment, on n'est pas forcément dans le répertoire où se trouve le programme python, aussi dans ce cas on peut utiliser un chemin « absolu », c'est-à-dire à partir de la racine des noms de fichiers, comme par exemple

```
$ python3 /le/chemin/jusqu/a/mon_module.py
```

Sauf que c'est assez malcommode, et cela devient vite pénible à la longue.

10.1.2 La solution

Sur linux et MacOS, il existe une astuce utile pour simplifier cela. Voyons comment s'y prendre, avec par exemple le programme `fibonacci.py` que vous pouvez [télécharger ici](#) (nous verrons ce code en détail dans les deux prochains compléments). Commencez par sauver ce code sur votre ordinateur dans un fichier qui s'appelle, bien entendu, `fibonacci.py`.

On commence par éditer le tout début du fichier pour lui ajouter une **première ligne**

```
#!/usr/bin/env python3

## La suite de Fibonacci (Suite)
...etc...
```

Cette première ligne s'appelle un **Shebang** dans le jargon Unix. Unix stipule que le Shebang doit être en **première position** dans le fichier.

Ensuite on rajoute au fichier, depuis le terminal, le caractère exécutable comme ceci

```
$ pwd
/le/chemin/jusqu/a/
```

```
$ chmod +x fibonacci.py
```

À partir de là, vous pouvez utiliser le fichier `fibonacci.py` comme une commande, sans avoir à mentionner `python3`, qui sera invoqué au travers du shebang.

```
$ /le/chemin/jusqu/a/fibonacci.py 20
fibonacci(20) = 10946
```

Et donc vous pouvez aussi le déplacer dans un répertoire qui est dans votre variable `PATH`, et le rendre ainsi accessible depuis n'importe quel répertoire en faisant simplement

```
$ export PATH=/le/chemin/jusqu/a:$PATH
```

```
$ cd /tmp
$ fibonacci.py 20
fibonacci(20) = 10946
```

Remarque : tout ceci fonctionne très bien tant que votre point d'entrée - ici `fibonacci.py` - n'utilise que des modules standards. Dans le cas où le point d'entrée vient avec au moins un module, il est également nécessaire d'installer ces modules quelque part, et d'indiquer au point d'entrée comment les trouver, nous y reviendrons dans la semaine où nous parlerons des modules.

CHAPITRE 11

Dessiner un carré

11.1 Exercice - niveau intermédiaire

Voici un tout petit programme qui dessine un carré.

Il utilise le module `turtle`, conçu précisément à des fins pédagogiques. Pour des raisons techniques, le module `turtle` n'est **pas disponible** au travers de la plateforme FUN.

Il est donc inutile d'essayer d'exécuter ce programme depuis le notebook. L'objectif de cet exercice est plutôt de vous entraîner à télécharger ce programme en utilisant le menu « *File -> Download as -> python* », puis à le charger dans votre IDLE pour l'exécuter sur votre machine.

Attention également à sauver le programme téléchargé **sous un autre nom** que `turtle.py`, car sinon vous allez empêcher python de trouver le module standard `turtle`; appelez-le par exemple `turtle_basic.py`.

```
# on a besoin du module turtle
import turtle
```

On commence par définir une fonction qui dessine un carré de côté `length`

```
def square(length):
    "have the turtle draw a square of side <length>"
    for side in range(4):
        turtle.forward(length)
        turtle.left(90)
```

Maintenant on commence par initialiser la tortue

```
turtle.reset()
```

On peut alors dessiner notre carré

```
square(200)
```

Et pour finir on attend que l'utilisateur clique dans la fenêtre de la tortue, et alors on termine

```
turtle.exitonclick()
```

11.2 Exercice - niveau avancé

Naturellement vous pouvez vous amuser à modifier ce code pour dessiner des choses un peu plus amusantes.

Dans ce cas, commencez par chercher « *module python turtle* » dans votre moteur de recherche favori, pour localiser la documentation du module `turtle` <<https://docs.python.org/3/library/turtle.html>>‘__.

Vous trouverez quelques exemples pour commencer ici :

- `turtle_multi_squares.py` pour dessiner des carrés à l’emplacement de la souris en utilisant plusieurs tortues ;
- `turtle_fractal.py` pour dessiner une fractale simple ;
- `turtle_fractal_reglable.py` une variation sur la fractale, plus paramétrable.

12.1 Complément - niveau basique

Revenons sur les noms de variables autorisés ou non.

Les noms les plus simples sont constitués de lettres. Par exemple

```
factoriel = 1
```

On peut utiliser aussi les majuscules, mais attention cela définit une variable différente. Ainsi

```
Factoriel = 100  
factoriel == Factoriel
```

```
False
```

Le signe == permet de tester si deux variables ont la même valeur. Si les variables ont la même valeur, le test retournera True, et False sinon. On y reviendra bien entendu.

12.1.1 Conventions habituelles

En règle générale, on utilise **uniquement des minuscules** pour désigner les variables simples (ainsi d'ailleurs que pour les noms de fonctions), les majuscules sont réservées en principe pour d'autres sortes de variables, comme les noms de classe, que nous verrons ultérieurement.

Notons, qu'il s'agit uniquement d'une convention, ceci n'est pas imposé par le langage lui-même.

Pour des raisons de lisibilité, il est également possible d'utiliser le tiret bas _ dans les noms de variables. On préférera ainsi

```
age_moyen = 75 # oui
```

plutôt que ceci (bien qu'autorisé par le langage)

```
AgeMoyen = 75 # autorisé, mais non
```

On peut également utiliser des chiffres dans les noms de variables comme par exemple

```
age_moyen_dept75 = 80
```

avec la restriction toutefois que le premier caractère ne peut pas être un chiffre, cette affectation est donc refusée

```
75_age_moyen = 80 # erreur de syntaxe
```

```
File "<ipython-input-6-823fed77034a>", line 1
  75_age_moyen = 80 # erreur de syntaxe
    ^
SyntaxError: invalid token
```

12.1.2 Le tiret bas comme premier caractère

Il est par contre, possible de faire commencer un nom de variable par un tiret bas comme premier caractère ; toutefois, à ce stade, nous vous déconseillons d'utiliser cette pratique qui est réservée à des conventions de nommage bien spécifiques.

```
_autorise_mais_deconseille = 'Voir le PEP 008'
```

Et en tous cas, il est **fortement déconseillé** d'utiliser des noms de la forme `__variable__` qui sont réservés au langage. Nous reviendrons sur ce point dans le futur, mais regardez par exemple cette variable que nous n'avons définie nulle part mais qui pourtant existe bel et bien

```
__name__ # ne définissez pas vous-même de variables de ce genre
```

```
'__main__'
```

12.1.3 Ponctuation

Dans la plage des caractères ASCII, il n'est **pas possible** d'utiliser d'autres caractères que les caractères alpha-numériques et le tiret bas. Notamment le tiret haut – est interprété comme l'opération de soustraction. Attention donc à cette erreur fréquente

```
age-moyen = 75 # erreur : en fait python lit 'age - moyen = 75'
```

```
File "<ipython-input-9-137d4530529a>", line 1
  age-moyen = 75 # erreur : en fait python lit 'age - moyen = 75'
                                     ^
SyntaxError: can't assign to operator
```

12.1.4 Caractères exotiques

En python-3, il est maintenant aussi possible d'utiliser des caractères Unicode dans les identificateurs

```
# les caractères accentués sont permis
nom_élève = "Jules Maigret"
```

```
# ainsi que l'alphabet grec
from math import cos, pi as
θ = / 4
cos(θ)
```

```
0.7071067811865476
```

Tous les caractères Unicode ne sont pas permis - heureusement car cela serait source de confusion. Nous citons dans les références les documents qui précisent quels sont exactement les caractères autorisés.

```
# ce caractère n'est pas autorisé
# il est considéré comme un signe mathématique (produit)
= 10
```

```
File "<ipython-input-12-7c140e06a1e3>", line 3
    = 10
    ^
SyntaxError: invalid character in identifier
```

Conseil

Il est **très vivement** recommandé :

- tout d’abord de coder **en anglais**,
- ensuite de **ne pas** définir des identificateurs avec des caractères non ASCII, dans toute la mesure du possible,
- enfin dès que vous mettez un caractère exotique dans votre code qui n’est pas supporté par UTF-8, vous devez bien spécifier l’encodage utilisé pour votre fichier source ; nous y reviendrons en deuxième semaine.

12.1.5 Pour en savoir plus

Pour les esprits curieux, Guido van Rossum, le fondateur de python, est le co-auteur d’un document qui décrit les conventions de codage à utiliser dans les bibliothèques python standard. Ces règles sont plus restrictives que ce que le langage permet de faire, mais constituent une lecture intéressante si vous projetez d’écrire beaucoup de python.

Voir dans le PEP 008 la [section consacrée aux règles de nommage](#) - (en anglais)

Voir enfin, au sujet des caractères exotiques dans les identificateurs :

- <https://www.python.org/dev/peps/pep-3131/> qui définit les caractères exotiques autorisés, et qui repose à son tour sur
- <http://www.unicode.org/reports/tr31/> (très technique !)

Les mots-clés de python

13.1 En Python

13.1.1 Mots réservés

Il existe en python certains mots spéciaux, qu'on appelle des mots-clés, ou *keywords* en anglais, qui sont réservés et **ne peuvent pas être utilisés** comme identifiants, c'est-à-dire comme un nom de variable.

C'est le cas par exemple pour l'instruction `if`, que nous verrons prochainement, qui permet bien entendu d'exécuter tel ou tel code selon le résultat d'un test.

```
variable = 15
if variable <= 10:
    print("au dessus de la moyenne")
else:
    print("en dessous")
```

```
en dessous
```

À cause de la présence de cette instruction dans le langage, il n'est pas autorisé d'appeler une variable `if`.

```
# interdit, if est un mot-clé
if = 1
```

```
File "<ipython-input-2-fl6082c36546>", line 2
    if = 1
    ^
SyntaxError: invalid syntax
```

13.1.2 Liste complète

Voici la liste complète des mots-clés :

| | | | | |
|--------------|----------|---------|-----------------|--------|
| False | class | finally | is | return |
| None | continue | for | lambda | try |
| True | def | from | nonlocal | while |
| and | del | global | not | with |
| as | elif | if | or | yield |
| assert | else | import | pass | |
| break | except | in | raise | |

Nous avons indiqué en gras les nouveautés par rapport à python-2 (sachant que réciproquement `exec` et `print` ont perdu leur statut de mot-clé depuis python-2, ce sont maintenant des fonctions).

Il vous faudra donc y prêter attention, surtout au début, mais avec un tout petit peu d'habitude vous saurez rapidement les éviter.

Vous remarquerez aussi que tous les bons éditeurs de texte supportant du code Python vont colorer les mots-clés différemment des variables. Par exemple, IDLE colorie les mots-clés en orange, vous pouvez donc très facilement vous rendre compte que vous allez, par erreur, en utiliser un comme nom de variable.

Cette fonctionnalité de *coloration syntaxique* permet d'identifier d'un coup d'oeil (grâce à un code de couleur) le rôle des différents éléments de votre code (variable, mots-clés, etc.) D'une manière générale, nous vous déconseillons fortement d'utiliser un éditeur de texte qui n'offre pas cette fonctionnalité de coloration syntaxique.

13.1.3 Pour en savoir plus

On peut se reporter à cette page

https://docs.python.org/3/reference/lexical_analysis.html#keywords

Un peu de calcul sur les types

14.1 Complément - niveau basique

14.1.1 La fonction `type`

Nous avons vu dans la vidéo que chaque objet possède un type. On peut très simplement accéder au type d'un objet en appelant une fonction *built-in*, c'est-à-dire prédéfinie dans python, qui s'appelle, eh bien oui, `type`.

On l'utilise tout simplement comme ceci

```
type(1)
```

```
int
```

```
type('spam')
```

```
str
```

Cette fonction est assez peu utilisée par les programmeurs expérimentés, mais va nous être utile à bien comprendre le langage, notamment pour manipuler les valeurs numériques.

14.1.2 Types, variables et objets

On a vu également que le type est attaché à **l'objet** et non à la variable.

```
x = 1  
type(x)
```

```
int
```

```
# la variable x peut référencer un objet de n'importe quel type  
  
x = [1, 2, 3]  
type(x)
```

```
list
```

14.2 Complément - niveau avancé

14.2.1 La fonction `isinstance`

Une autre fonction prédéfinie, voisine de `type` mais plus utile dans la pratique, est la fonction `isinstance` qui permet de savoir si un objet est d'un type donné. Par exemple

```
isinstance(23, int)
```

```
True
```

À la vue de ce seul exemple, on pourrait penser que `isinstance` est presque identique à `type` ; en réalité elle est un peu plus élaborée, notamment pour la programmation objet et l'héritage, nous aurons l'occasion d'y revenir.

On remarque ici en passant que la variable `int` est connue de python alors que nous ne l'avons pas définie. Il s'agit d'une variable prédéfinie, qui désigne le type des entiers, que nous étudierons très bientôt.

Pour conclure sur `isinstance`, cette fonction est utile en pratique précisément parce que python est à typage dynamique. Aussi il est souvent utile de s'assurer qu'une variable passée à une fonction est du (ou des) type(s) attendu(s), puisque contrairement à un langage typé statiquement comme C++, on n'a aucune garantie de ce genre à l'exécution. À nouveau, nous aurons l'occasion de revenir sur ce point.

15.1 Complément - niveau basique

L'objet de ce complément est de vous montrer qu'avec python vous n'avez pas à vous préoccuper de la mémoire. Pour expliquer la notion de gestion de la mémoire, il nous faut donner un certain nombre de détails sur d'autres langages comme C et C++. Si vous souhaitez suivre ce cours à un niveau basique vous pouvez ignorer ce complément et seulement retenir que python se charge de tout pour vous :)

15.2 Complément - niveau intermédiaire

15.2.1 Langages de bas niveau

Dans un langage traditionnel de bas niveau comme C ou C++, le programmeur est en charge de l'allocation - et donc de la libération - de la mémoire.

Ce qui signifie que, sauf pour les valeurs stockées dans la pile, le programmeur est amené

- à réclamer de la mémoire à l'OS en appelant explicitement `malloc` (C) ou `new` (C++)
- et réciproquement à rendre cette mémoire à l'OS lorsqu'elle n'est plus utilisée, en appelant `free` (C) ou `delete` (C++).

Avec ce genre de langage, la gestion de la mémoire est un aspect important de la programmation. Ce modèle offre une grande flexibilité, mais au prix d'un coût élevé en termes de vitesse de développement.

En effet, il est assez facile d'oublier de libérer la mémoire après usage, ce qui peut conduire à épuiser les ressources disponibles. À l'inverse, utiliser une zone mémoire non allouée peut conduire à des bugs très difficiles à localiser et à des problèmes de sécurité majeurs. Notons qu'une grande partie des attaques en informatique reposent sur l'exploitation d'erreurs de gestion de la mémoire.

15.2.2 Langages de haut niveau

Pour toutes ces raisons, avec un langage de plus haut niveau comme python, le programmeur est libéré de cet aspect de la programmation.

Pour anticiper un peu sur le cours des semaines suivantes, voici ce que vous pouvez garder en tête s'agissant de la gestion mémoire en python.

- Vous créez vos objets au fur et à mesure de vos besoins.
- Vous n'avez pas besoin de les libérer explicitement, le « *Garbage Collector* » de python va s'en charger pour recycler la mémoire lorsque c'est possible.
- Python a tendance à être assez gourmand en mémoire, comparé à un langage de bas niveau, car tout est objet et chaque objet est assorti de *méta-informations* qui occupent une place non négligeable. Par exemple, chaque objet possède au minimum
 - une référence vers son type - c'est le prix du typage dynamique,
 - un compteur de références - le nombre d'autres valeurs (variables ou objets) qui pointent vers l'objet, cette information est notamment utilisée, précisément, par le *Garbage Collector* pour déterminer si la mémoire utilisée par un objet peut être libérée ou non.
- Un certain nombre de types prédéfinis et non mutables sont implémentés en python comme des *singletons*, c'est-à-dire qu'un seul objet est créé et partagé, c'est le cas par exemple pour les petits entiers et les chaînes de caractères, on en reparlera.
- Lorsqu'on implémente une classe, il est possible de lui conférer cette caractéristique de singleton, de manière à optimiser la mémoire nécessaire pour exécuter un programme.

Typages statique et dynamique

16.1 Complément - niveau intermédiaire

Parmi les langages typés, on distingue les langages à typage statique et à typage dynamique. Ce notebook tente d'éclaircir ces notions pour ceux qui n'y sont pas familiers.

16.1.1 Typage statique

À une extrémité du spectre, on trouve les langages compilés, dits à typage statique, comme par exemple C/C++.

En C on écrira, par exemple, une version simpliste de la fonction factoriel comme ceci

```
#include <stdio.h>

int factoriel (int n) {
    int result = 1;
    int loop;
    for (loop = 1; loop <= n; loop++) {
        result *= loop;
    }
    return result;
}
```

Comme vous pouvez le voir - ou le deviner - toutes les **variables** utilisées ici (comme par exemple `n`, `result` et `loop`) sont typées.

- On doit appeler `factoriel` avec un argument `n` qui doit être un entier (`int` est le nom du type entier).
- Les variables internes `result` et `loop` sont de type entier.
- `factoriel` retourne une valeur de type entier.

Ces informations de type ont essentiellement trois fonctions.

- En premier lieu, elles sont nécessaires au compilateur. En C si le programmeur ne précisait pas que `result` est de type entier, le compilateur n'aurait pas suffisamment d'éléments pour générer le code assembleur correspondant.
- En contrepartie, le programmeur a un contrôle très fin de l'usage qu'il fait de la mémoire et du hardware. Il peut choisir d'utiliser un entier sur 32 ou 64 bits, signé ou pas, ou construire avec `struct` et `union` un arrangement de ses données.
- Enfin, et surtout, ces informations de type permettent de faire un contrôle *a priori* de la validité du programme, par exemple, si à un autre endroit dans le code on trouve

```
int main (int argc, char *argv[]) {
    /* le premier argument de la ligne de commande est argv[1] */
    char *input = argv[1];
    /* calculer son factoriel et afficher le resultat */
    printf ("Factoriel (%s) = %d\n", input, factoriel(input));
    /*                                     ^^^                                     */
    /* ici on appelle factoriel avec une entree 'chaîne de caractere' */
    return 0;
}
```

alors le compilateur va remarquer qu'on essaie d'appeler `factoriel` avec comme argument `input` qui, pour faire simple, est une chaîne de caractères et comme `factoriel` s'attend à recevoir un entier, ce programme n'a aucune chance de fonctionner.

On parle alors de **typage statique**, en ce sens que chaque **variable** a exactement un type qui est défini par le programmeur une bonne fois pour toutes.

C'est ce qu'on appelle le **contrôle de type**, ou *type-checking* en anglais. Si on ignore le point sur le contrôle fin de la mémoire, qui n'est pas crucial à notre sujet, ce modèle de contrôle de type présente :

- l'**inconvenient** de demander davantage au programmeur (je fais abstraction, à ce stade et pour simplifier, de langages à inférence de types comme ML et Haskell)
- et l'**avantage** de permettre un contrôle étendu, et surtout précoce (avant même de l'exécuter), de la bonne correction du programme.

Cela étant dit, le typage statique en C n'empêche pas le programmeur débutant d'essayer d'écrire dans la mémoire à partir d'un pointeur `NULL` - et le programme de s'interrompre brutalement. Il faut être conscient des limites du typage statique.

16.1.2 Typage dynamique

À l'autre bout du spectre, on trouve des langages comme, eh bien, python.

Pour comprendre cette notion de typage dynamique, regardons la fonction suivante `somme`.

```
def somme(*largs):
    "retourne la somme de tous ses arguments"
    if not largs:
        return 0
    elif len(largs) == 1:
        return largs[0]
    else:
        result = largs[0] + largs[1]
        for i in range(2, len(largs)):
            result += largs[i]
        return result
```

Naturellement, vous n'êtes pas à ce stade en mesure de comprendre le fonctionnement intime de la fonction. Mais vous pouvez tout de même l'utiliser.


```
somme(12, 14, 300)
```

```
326
```

```
l1 = ['a', 'b', 'c']
l2 = [0, 20, 30]
l3 = ['spam', 'eggs']
somme(l1, l2, l3)
```

```
['a', 'b', 'c', 0, 20, 30, 'spam', 'eggs']
```

Vous pouvez donc constater que `somme` peut fonctionner avec des objets de types différents. En fait, telle qu'elle est écrite, elle va fonctionner s'il est possible de faire `+` entre ses arguments. Ainsi par exemple on pourrait même faire

```
# python sait faire + entre deux chaînes de caractères
somme('abc', 'def')
```

```
'abcdef'
```

Mais par contre on ne pourrait pas faire

```
# ceci va déclencher une exception à run-time
somme(12, [1, 2, 3])
```

```
-----
TypeError                                Traceback (most recent call last)

<ipython-input-6-8ca5fddecf76> in <module>()
      1 # ceci va déclencher une exception à run-time
----> 2 somme(12, [1, 2, 3])

<ipython-input-2-359eb67dbb5d> in somme(*largs)
      6         return largs[0]
      7     else:
----> 8         result = largs[0] + largs[1]
      9         for i in range(2, len(largs)):
     10             result += largs[i]

TypeError: unsupported operand type(s) for +: 'int' and 'list'
```

Il est utile de remarquer que le typage de python, qui existe bel et bien comme on le verra, est qualifié de dynamique parce que le type est attaché à **un objet** et non à la variable qui le référence. On aura bien entendu l'occasion d'approfondir tout ça dans le cours.

En python, on fait souvent référence au typage sous l'appellation *duck typing*, de manière imagée

If it looks like a duck and quacks like a duck, it's a duck

On voit qu'on se trouve dans une situation très différente de celle du programmeur C/C++, en ce sens que :

- à l'écriture du programme, il n'y a aucun des surcoûts qu'on trouve avec C ou C++ en terme de définition de type,
- aucun contrôle de type n'est effectué *a priori* par le langage au moment de la définition de la fonction `somme`,

- par contre au moment de l'exécution, s'il s'avère qu'on tente de faire une somme entre deux types qui ne peuvent pas être additionnés, comme ci-dessus avec un entier et une liste, le programme ne pourra pas se dérouler correctement.

Il y a deux points de vue vis-à-vis de la question du typage.

Les gens habitués au *typage statique* se plaignent du typage dynamique en disant qu'on peut écrire des programmes faux et qu'on s'en rend compte trop tard - à l'exécution.

À l'inverse les gens habitués au *typage dynamique* font valoir que le typage statique est très partiel, par exemple, en C si on essaie d'écrire au bout d'un pointeur nul, l'OS ne le permet pas et le programme sort tout aussi brutalement.

Bref, selon le point de vue, le typage dynamique est vécu comme un inconvénient (pas assez de bonnes propriétés détectées par le langage) ou comme un avantage (pas besoin de passer du temps à déclarer le type des variables, ni à faire des conversions pour satisfaire le compilateur). Vous remarquerez cependant qu'à l'usage, en terme de vitesse de développement, les inconvénients du typage dynamique sont très largement compensés par ses avantages.

16.1.3 Type hints

Signalons enfin que depuis python-3.5, il est **possible** d'ajouter des annotations de type, pour expliciter les suppositions qui sont faites par le programmeur pour le bon fonctionnement du code.

Nous aurons là encore l'occasion de détailler ce point dans le cours, signalons simplement que ces annotations sont totalement optionnelles, et que même lorsqu'elles sont présentes elles ne sont pas utilisées à runtime par l'interpréteur. L'idée est plutôt de permettre à de outils externes, *comme par exemple* *'mypy'* <<http://www.mypy-lang.org>>__, d'effectuer des contrôles plus poussés concernant la correction du programme.

Utiliser python comme une calculette

17.1 Dans l'interpréteur

Lorsque vous démarrez l'interprète python, vous disposez en fait d'une calculette, par exemple, vous pouvez taper

```
20 * 60
```

```
1200
```

Les règles de **priorité** entre les opérateurs sont habituelles, les produits et divisions sont évalués en premier, ensuite les sommes et soustractions

```
2 * 30 + 10 * 5
```

```
110
```

De manière générale, il est recommandé de bien parenthéser ses expressions. De plus, les parenthèses facilitent la lecture d'expressions complexes.

Par exemple, il vaut mieux écrire ce qui suit, qui est équivalent mais plus lisible

```
(2 * 30) + (10 * 5)
```

```
110
```

Attention, en python la division / est une division naturelle

```
48 / 5
```

```
9.6
```

Rappelez-vous des opérateurs suivants qui sont très pratiques

| code | opération |
|------|-----------|
| // | quotient |
| % | modulo |
| ** | puissance |

```
# calculer un quotient
48 // 5
```

```
9
```

```
# modulo (le reste de la division par)
48 % 5
```

```
3
```

```
# puissance
2 ** 10
```

```
1024
```

Vous pouvez facilement faire aussi des calculs sur les complexes. Souvenez-vous seulement que la constante complexe que nous notons en français *i* se note en python *j*, ce choix a été fait par le BDFL - alias Guido van Rossum - pour des raisons de lisibilité

```
# multiplication de deux nombres complexes
(2 + 3j) * 2.5j
```

```
(-7.5+5j)
```

Aussi, pour entrer ce nombre complexe *j*, il faut toujours le faire précéder d'un nombre, donc ne pas entrer simplement *j* (qui serait compris comme un nom de variable, nous allons voir ça tout de suite) mais plutôt *1j* ou encore *1. j*, comme ceci

```
1j * 1.j
```

```
(-1+0j)
```

17.1.1 Utiliser des variables

Il peut être utile de stocker un résultat qui sera utilisé plus tard, ou de définir une valeur constante. Pour cela on utilise tout simplement une affectation comme ceci

```
# pour définir une variable il suffit de lui assigner une valeur
largeur = 5
```

```
# une fois la variable définie, on peut l'utiliser, ici comme un nombre
largeur * 20
```

```
100
```

```
# après quoi bien sûr la variable reste inchangée
largeur * 10
```

```
50
```

Pour les symboles mathématiques, on peut utiliser la même technique

```
# pour définir un réel, on utilise le point au lieu d'une virgule en français
pi = 3.14159
2 * pi * 10
```

```
62.8318
```

Pour les valeurs spéciales comme π , on peut utiliser les valeurs prédéfinies par la librairie mathématique de python. En anticipant un peu sur la notion d'importation que nous approfondirons plus tard, on peut écrire

```
from math import e, pi
```

Et ainsi imprimer les racines troisièmes de l'unité par la formule

$$r_n = e^{2i\pi \frac{n}{3}}, \text{ pour } n \in \{0, 1, 2\}$$

```
n = 0
print("n=", n, "racine = ", e**((2.j*pi*n)/3))
n = 1
print("n=", n, "racine = ", e**((2.j*pi*n)/3))
n = 2
print("n=", n, "racine = ", e**((2.j*pi*n)/3))
```

```
n= 0 racine = (1+0j)
n= 1 racine = (-0.4999999999999998+0.8660254037844387j)
n= 2 racine = (-0.5000000000000004-0.8660254037844384j)
```

Remarque : bien entendu il sera possible de faire ceci plus simplement lorsque nous aurons vu les boucles `for`.

17.1.2 Les types

Ce qui change par rapport à une calculatrice standard est le fait que les valeurs sont typées. Pour illustrer les trois types de nombres que nous avons vus jusqu'ici

```
# le type entier s'appelle 'int'
type(3)
```

```
int
```

```
# le type flottant s'appelle 'float'
type(3.5)
```

```
float
```

```
# le type complexe s'appelle 'complex'
type(1j)
```

```
complex
```

17.1.3 Chaînes de caractères

On a également rapidement besoin de chaînes de caractères, on les étudiera bientôt en détails, mais en guise d'avant-goût

```
chaine = "Bonjour le monde !"
print(chaine)
```

```
Bonjour le monde !
```

17.1.4 Conversions

Il est parfois nécessaire de convertir une donnée d'un type dans un autre. Par exemple on peut demander à l'utilisateur d'entrer une valeur au clavier grâce à la fonction `input`, comme ceci

```
reponse = input("quel est votre âge ? ")
```

```
quel est votre âge ? 45
```

```
# vous avez entré la chaîne suivante
print(reponse)
```

```
45
```

```
# ici reponse est une variable, et son contenu est de type chaîne de caractères
type(reponse)
```

```
str
```

Maintenant je veux faire des calculs sur votre âge, par exemple le multiplier par 2. Si je m'y prends naïvement, ça donne ceci

```
# multiplier une chaîne de caractères par deux ne fait pas ce qu'on veut,
# nous verrons plus tard que ça fait une concaténation
2 * reponse
```

```
'4545'
```

C'est pourquoi il me faut ici d'abord **convertir** la (valeur de la) variable `reponse` en un entier, que je peux ensuite doubler (assurez-vous d'avoir bien entré ci-dessus une valeur qui correspond à un nombre entier)

```
# reponse est une chaîne
# je la convertis en entier en appelant la fonction int()
age = int(reponse)
type(age)
```

```
int
```

```
# que je peux maintenant multiplier par 2
2 * age
```

```
90
```

Ou si on préfère, en une seule fois

```
print("le double de votre age est", 2*int(reponse))
```

```
le double de votre age est 90
```

17.1.5 Conversions - suite

De manière plus générale, pour convertir un objet en un entier, un flottant, ou une chaîne de caractères, on peut simplement appeler une fonction built-in qui porte le même nom que le type cible.

| Type | Fonction |
|----------|----------|
| Entier | int |
| Flottant | float |
| Complexe | complex |
| Chaîne | str |

Ainsi dans l'exemple précédent, `int(reponse)` représente la conversion de `reponse` en entier.

On a illustré cette même technique dans les exemples suivants.

```
# dans l'autre sens, si j'ai un entier
a = 2345
```

```
# je peux facilement le traduire en chaîne de caractères
str(2345)
```

```
'2345'
```

```
# ou en complexe
complex(2345)
```

```
(2345+0j)
```

Nous verrons plus tard que ceci se généralise à tous les types de python, pour convertir un objet `x` en un type `bidule`, on appelle `bidule(x)`. On y reviendra, bien entendu.

17.1.6 Grands nombres

Comme les entiers sont de précision illimitée, on peut améliorer leur lisibilité en insérant des caractères `_` qui sont simplement ignorés à l'exécution.

```
tres_grand_nombre = 23_456_789_012_345

tres_grand_nombre
```

```
23456789012345
```

```
# ça marche aussi avec les flottants
123_456.789_012
```

```
123456.789012
```

17.1.7 Entiers et bases

Les calculatrices scientifiques permettent habituellement d'entrer les entiers dans d'autres bases que la base 10.

En python, on peut aussi entrer un entier sous forme binaire comme ceci

```
deux_cents = 0b11001000 ; print(deux_cents)
```

```
200
```

Ou encore sous forme octale (en base 8) comme ceci

```
deux_cents = 0o310 ; print(deux_cents)
```

```
200
```

Ou enfin encore en hexadecimal (base 16) comme ceci

```
deux_cents = 0xc8 ; print(deux_cents)
```

```
200
```

Pour d'autres bases, on peut utiliser la fonction de conversion `int` en lui passant un argument supplémentaire.

```
deux_cents = int('3020', 4) ; print(deux_cents)
```

```
200
```

17.1.8 Fonctions mathématiques

python fournit naturellement un ensemble très complet d'opérateurs mathématiques pour les fonctions exponentielles, trigonométriques et autres, mais leur utilisation ne nous est pas encore accessible à ce stade et nous les verrons ultérieurement.

Affectations & Opérations (à la +=)

18.1 Complément - niveau intermédiaire

Il existe en python toute une famille d'opérateurs dérivés de l'affectation qui permettent de faire en une fois une opération et une affectation. En voici quelques exemples.

18.1.1 Incrémentation

On peut facilement augmenter la valeur d'une variable numérique comme ceci

```
entier = 10

entier += 2
print('entier', entier)
```

```
entier 12
```

Comme on le devine peut-être, ceci est équivalent à

```
entier = 10

entier = entier + 2
print('entier', entier)
```

```
entier 12
```

18.1.2 Autres opérateurs courants

Cette forme, qui combine opération sur une variable et réaffectation du résultat à la même variable, est disponible avec tous les opérateurs courants.

```
entier -= 4
print('après décrémentation', entier)
entier *= 2
print('après doublement', entier)
```

```
entier /= 2
print('mis à moitié', entier)
```

```
après décréement 8
après doublement 16
mis à moitié 8.0
```

18.1.3 Types non numériques

En réalité cette construction est disponible sur tous les types qui supportent l'opérateur en question. Par exemple, les listes (que nous verrons bientôt) peuvent être additionnées entre elles.

```
liste = [0, 3, 5]
print('liste', liste)

liste += ['a', 'b']
print('après ajout', liste)
```

```
liste [0, 3, 5]
après ajout [0, 3, 5, 'a', 'b']
```

Beaucoup de types supportent l'opérateur +, qui est sans doute de loin celui qui est le plus utilisé avec cette construction.

18.1.4 Opérateurs plus abscons

Signalons enfin qu'on trouve cette construction aussi avec d'autres opérateurs moins fréquents, par exemple

```
entier = 2
print('entier:', entier)
entier **= 10
print('à la puissance dix:', entier)
entier %= 5
print('modulo 5:', entier)
entier <<= 2
print('double décalage gauche:', entier)
```

```
entier: 2
à la puissance dix: 1024
modulo 5: 4
double décalage gauche: 16
```

Notions sur la précision des calculs flottants

19.1 Complément - niveau avancé

19.1.1 Le problème

Comme pour les entiers, les calculs sur les flottants sont, naturellement, réalisés par le processeur. Cependant contrairement au cas des entiers où les calculs sont toujours exacts, les flottants posent un problème de précision. Cela n'est pas propre au langage python, mais est dû à la technique de codage des nombres flottants sous forme binaire.

Voyons tout d'abord comment se matérialise le problème.

```
0.2 + 0.4
```

```
0.6000000000000001
```

Il faut retenir que lorsqu'on écrit un nombre flottant sous forme décimale, la valeur utilisée en mémoire pour représenter ce nombre, parce que cette valeur est codée en binaire, ne représente **pas toujours exactement** le nombre entré.

```
# du coup cette expression est fausse, à cause de l'erreur d'arrondi
0.3 - 0.1 == 0.2
```

```
False
```

Aussi comme on le voit les différentes erreurs d'arrondi qui se produisent à chaque étape du calcul s'accumulent et produisent un résultat parfois surprenant. De nouveau, ce problème n'est pas spécifique à python, il existe pour tous les langages et il est bien connu des numériciens.

Dans une grande majorité des cas, ces erreurs d'arrondi ne sont pas pénalisantes. Il faut toutefois en être conscient car cela peut expliquer des comportements curieux.

19.1.2 Une solution : penser en termes de nombres rationnels

Tout d'abord si votre problème se pose bien en termes de nombres rationnels, il est alors tout à fait possible de le résoudre avec exactitude.

Alors qu'il n'est pas possible d'écrire exactement $3/10$ en base 2, ni d'ailleurs $1/3$ en base 10, on peut représenter **exactement** ces nombres dès lors qu'on les considère comme des fractions et qu'on les encode avec deux nombres entiers.

Python fournit en standard le module `fractions` qui permet de résoudre le problème. Voici comment on pourrait l'utiliser pour vérifier, cette fois avec succès, que $0.3 - 0.1$ vaut bien 0.2 . Ce code anticipe sur l'utilisation des modules et des classes en Python, ici nous créons des objets de type `Fraction`

```
# on importe le module fractions, qui lui-même définit le symbole Fraction
from fractions import Fraction

# et cette fois, les calculs sont exacts, et l'expression retourne bien True
Fraction(3, 10) - Fraction(1, 10) == Fraction(2, 10)
```

```
True
```

Ou encore d'ailleurs, équivalent et plus lisible

```
Fraction('0.3') - Fraction('0.1') == Fraction('2/10')
```

```
True
```

19.1.3 Une autre solution : le module decimal

Si par contre vous ne manipulez pas des nombres rationnels et que du coup la représentation sous forme de fractions ne peut pas convenir dans votre cas, signalons l'existence du module standard `decimal` qui offre des fonctionnalités très voisines du type `float`, tout en éliminant la plupart des inconvénients, au prix naturellement d'une consommation mémoire supérieure.

Pour reprendre l'exemple de départ, mais en utilisant le module `decimal`, on écrirait alors

```
from decimal import Decimal

Decimal('0.3') - Decimal('0.1') == Decimal('0.2')
```

```
True
```

19.1.4 Pour aller plus loin

Tous ces documents sont en anglais.

- [Un tutoriel sur les nombres flottants](#)
- [La documentation sur la classe Fraction](#)
- [La documentation sur la classe Decimal](#)

20.1 Compléments - niveau avancé

Les compléments ci-dessous expliquent des fonctions évoluées sur les entiers. Les débutants en programmation peuvent sans souci sauter cette partie en cas de difficultés.

20.1.1 Opérations logiques : ET &, OU | et OU exclusif ^

Il est possible aussi de faire des opérations *bit-à-bit* sur les nombres entiers. Le plus simple est de penser à l'écriture du nombre en base 2.

Considérons par exemple deux entiers constants dans cet exercice

```
x49 = 49
y81 = 81
```

Ce qui nous donne comme décomposition binaire

$$\begin{aligned} x_{49} &= 49 = 32 + 16 + 1 \rightarrow (0, 1, 1, 0, 0, 0, 1) \\ y_{81} &= 81 = 64 + 16 + 1 \rightarrow (1, 0, 1, 0, 0, 0, 1) \end{aligned}$$

Pour comprendre comment passer de $32 + 16 + 1$ à $(0, 1, 1, 0, 0, 0, 1)$ il suffit d'observer que

$$32 + 16 + 1 = 0 * 2^6 + 1 * 2^5 + 1 * 2^4 + 0 * 2^3 + 0 * 2^2 + 0 * 2^1 + 1 * 2^0$$

Et logique : opérateur &

L'opération logique & va faire un *et* logique bit à bit entre les opérandes, ainsi

```
x49 & y81
```

```
17
```

$$\begin{aligned} \text{Et en effet} \quad x_{49} &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\ y_{81} &\rightarrow (1, 0, 1, 0, 0, 0, 1) \\ x_{49} \& y_{81} &\rightarrow (0, 0, 1, 0, 0, 0, 1) \rightarrow 16 + 1 \rightarrow 17 \end{aligned}$$

Ou logique : opérateur |

De même, l'opérateur logique | fait simplement un *ou* logique, comme ceci

```
x49 | y81
```

```
113
```

On s'y retrouve parce que

$$\begin{aligned}
 x49 &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\
 y81 &\rightarrow (1, 0, 1, 0, 0, 0, 1) \\
 x49 | y81 &\rightarrow (1, 1, 1, 0, 0, 0, 1) \rightarrow 64 + 32 + 16 + 1 \rightarrow 113
 \end{aligned}$$

Ou exclusif : opérateur ^

Enfin on peut également faire la même opération à base de *ou exclusif* avec l'opérateur ^

```
x49 ^ y81
```

```
96
```

Je vous laisse le soin de décortiquer le calcul à titre d'exercice (le ou exclusif de deux bits est vrai si et seulement si exactement une des deux entrées est vraie).

20.1.2 Décalages

Un décalage à *gauche* de, par exemple, 4 positions, revient à décaler tout le champ de bits de 4 cases à gauche (les 4 nouveaux bits insérés sont toujours des 0). C'est donc équivalent à une multiplication par $2^4 = 16$

```
x49 << 4
```

```
784
```

$$\begin{aligned}
 x49 &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\
 x49 << 4 &\rightarrow (0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0) \rightarrow 512 + 256 + 16 \rightarrow 784
 \end{aligned}$$

De la même façon le décalage à droite de n revient à une division par 2^n (plus précisément, le quotient de la division)

```
x49 >> 4
```

```
3
```

$$\begin{aligned}
 x49 &\rightarrow (0, 1, 1, 0, 0, 0, 1) \\
 x49 >> 4 &\rightarrow (0, 0, 0, 0, 0, 1, 1) \rightarrow 2 + 1 \rightarrow 3
 \end{aligned}$$

20.1.3 Une astuce

On peut utiliser la fonction *built-in* `bin` pour calculer la représentation binaire d'un entier, attention la valeur de retour est une chaîne de caractères de type `str`

```
bin(x49)
```

```
'0b110001'
```

Dans l'autre sens, on peut aussi entrer un entier directement en base 2 comme ceci, ici comme on le voit `x49bis` est bien un entier

```
x49bis = 0b110001  
x49bis == x49
```

```
True
```

20.1.4 Pour en savoir plus

[Section de la documentation python](#)

Estimer le plus petit (grand) flottant

21.1 Exercice - niveau basique

21.1.1 Le plus petit flottant

En corollaire de la discussion sur la précision des flottants, il faut savoir que le système de codage en mémoire impose aussi une limite. Les réels très petits, ou très grands, ne peuvent plus être représentés de cette manière.

C'est notamment très gênant si vous implémentez un logiciel probabiliste, comme des graphes de Markov, où les probabilités d'occurrence de séquences très longues tendent très rapidement vers des valeurs extrêmement petites.

Le but de cet exercice est d'estimer la valeur du plus petit flottant qui peut être représenté comme un flottant. Pour vous aider, voici deux valeurs

10**-320

1e-320

10**-330

0.0

Comme on le voit, 10^{-320} est correctement imprimé, alors que 10^{-330} est, de manière erronée, rapporté comme étant nul.

Notes

- À ce stade du cours, pour estimer le plus petit flottant, procédez simplement par approximations successives.
- Sans utiliser de boucle, la précision que vous pourrez obtenir n'est que fonction de votre patience, ne dépassez pas 4 à 5 itérations successives :)
- Il est par contre pertinent d'utiliser une approche rationnelle pour déterminer l'itération suivante (par opposition à une approche « au petit bonheur »). Pour ceux qui ne connaissent pas, nous vous recommandons de vous documenter sur l'algorithme de **dichotomie**.

Voici quelques cellules de code vides, vous pouvez en créer d'autres si nécessaire, le plus simple étant de taper **Alt+Enter**, ou d'utiliser le menu « *Insert -> Insert Cell Below* »

21.1.2 Le plus grand flottant

La même limitation s'applique sur les grands nombres. Toutefois cela est un peu moins évident, car comme toujours il faut faire attention aux types

[illegible]

Ce qui se passe très bien car on j'ai utilisé un `int` pour l'exposant, dans ce premier cas python calcule le résultat comme un `int`, qui est un type qui n'a pas de limitation de précision (python utilise intelligemment autant de bits que nécessaire pour ce genre de calculs).

Par contre, si j'essaie de faire le même calcul avec un exposant flottant, python essaie cette fois de faire son calcul avec un flottant, et là on obtient une erreur

[illegible]

```
<ipython-input-7-d63f7e475389> in <module>()
----> 1 10**450.0
```

```
OverflowError: (34, 'Numerical result out of range')
```

On peut d'ailleurs remarquer que le comportement ici n'est pas extrêmement cohérent, car avec les petits nombres python nous a silencieusement transformé 10^{-330} en 0, alors que pour les grands nombres, il lève une exception (nous verrons les exceptions plus tard, mais vous pouvez dès maintenant remarquer que le comportement est différent dans les deux cas).

Quoi qu'il en soit, la limite pour les grands nombres se situe entre les deux valeurs 10^{300} et 10^{310} . On vous demande à nouveau d'estimer comme ci-dessus une valeur approchée du plus grand nombre qu'il est possible de représenter comme un flottant.

```
10**300.
```

```
1e+300
```

```
10**310.
```

```
-----
OverflowError                                Traceback (most recent call last)
<ipython-input-9-5d701e1fa38c> in <module>()
----> 1 10**310.

OverflowError: (34, 'Numerical result out of range')
```

```
10**309.
```

```
-----
OverflowError                                Traceback (most recent call last)
<ipython-input-18-f39c19dlb8b5> in <module>()
----> 1 10**309.

OverflowError: (34, 'Numerical result out of range')
```

21.2 Complément - niveau avancé

En fait, on peut accéder à ces valeurs minimales et maximales pour les flottants comme ceci

```
import sys
print(sys.float_info)
```

```
sys.float_info(max=1.7976931348623157e+308, max_exp=1024, max_10_exp=308, min=2.
↪2250738585072014e-308, min_exp=-1021, min_10_exp=-307, dig=15, mant_dig=53,
↪epsilon=2.220446049250313e-16, radix=2, rounds=1)
```

Et notamment, comme expliqué ici

```
print("Flottant minimum", sys.float_info.min)
print("Flottant maximum", sys.float_info.max)
```

```
Flottant minimum 2.2250738585072014e-308
Flottant maximum 1.7976931348623157e+308
```

Sauf que vous devez avoir trouvé un maximum voisin de cette valeur, mais le minimum observé expérimentalement ne correspond pas bien à cette valeur.

Pour ceux que cela intéresse, l'explication à cette apparente contradiction réside dans l'utilisation de **nombre**s dénormaux.

Caractères accentués

Ce complément expose quelques bases concernant les caractères accentués, et notamment les précautions à prendre pour pouvoir en insérer dans un programme python. Nous allons voir que cette question, assez scabreuse, dépasse très largement le cadre de python.

22.1 Complément - niveau basique

22.1.1 Un caractère ce n'est pas un octet

Avec Unicode, on a cassé le modèle *un caractère == un octet*. Aussi en python-3, lorsqu'il s'agit de manipuler des données provenant de diverses sources de données :

- le type `byte` est approprié si vous voulez charger en mémoire les données binaires brutes, sous forme d'octets donc,
- le type `str` est approprié pour représenter une chaîne de caractères - qui, à nouveau ne sont pas forcément des octets,
- on passe de l'un à l'autre de ces types par des opérations d'encodage et décodage, comme illustré ci-dessous,
- et pour **toutes** les opérations d'encodage et décodage, il est nécessaire de connaître l'encodage utilisé.

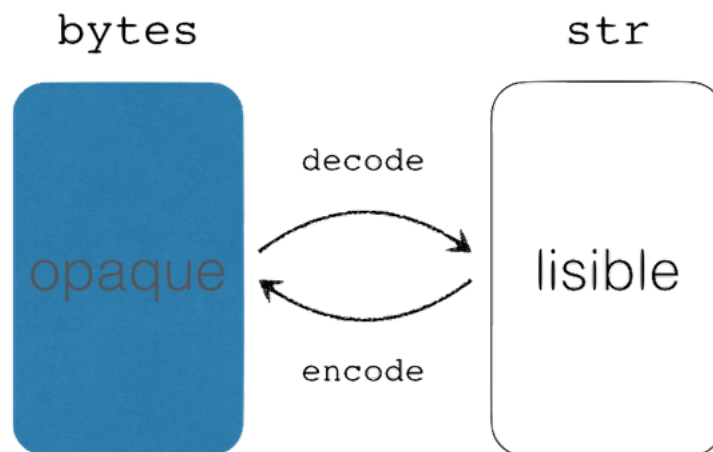


Fig. 22.1 – les types bytes et str

On peut appeler les méthodes `encode` et `decode` sans préciser l’encodage (dans ce cas python choisit l’encodage par défaut sur votre système). Cela dit, il est de loin préférable d’être explicite et de choisir son encodage. En cas de doute, il est recommandé de **spécifier explicitement** `utf-8`, qui se généralise au détriment d’encodages anciens comme `cp1242` (windows) et `iso-latin-*`, que de laisser le système hôte choisir pour vous.

22.1.2 Utilisation des accents et autres cédilles

Python-3 supporte Unicode par défaut. Vous pouvez donc, maintenant, utiliser sans aucun risque des accents ou des cédilles dans vos chaînes de caractères. Il faut cependant faire attention à deux choses.

- Python supporte Unicode, donc tous les caractères du monde, mais les ordinateurs n’ont pas forcément les polices de caractères nécessaires pour afficher ces caractères.
- Python permet d’utiliser des caractères Unicode pour les noms de variables, mais nous vous recommandons dans toute la mesure du possible d’écrire votre code en anglais, comme c’est le cas pour la quasi-totalité du code que vous serez amenés à utiliser sous forme de librairies.

Ainsi, il faut bien distinguer les chaînes de caractères qui doivent par nature être adaptées au langage des utilisateurs du programme, et le code source qui lui est destiné aux programmeurs et qui doit donc éviter d’utiliser autre chose que de l’anglais.

22.2 Complément - niveau intermédiaire

22.2.1 Accents et encodage

Où peut-on mettre des accents ?

Cela étant dit, si vous devez vraiment mettre des accents dans vos sources, voici ce qu’il faut savoir.

Noms de variables

- S’il n’était **pas possible en python-2** d’utiliser un caractère accentué dans un **nom de variable** (ou d’un identificateur au sens large), cela est à présent **permis en python-3** :

```
# pas recommandé, mais autorisé par le langage
nb_élèves = 12
```

- On peut même utiliser des symboles, comme par exemple

```
from math import cos, pi as Π
θ = Π / 4
cos(θ)
```

- Je vous recommande toutefois de **ne pas utiliser** cette possibilité, si vous n’êtes pas extrêmement familier avec les caractères Unicode.
- Enfin pour être exhaustif, sachez que seule une partie des caractères Unicode sont autorisés dans ce cadre, c’est heureux parce que les caractères comme, par exemple, l’espace non-sécable pourraient, s’ils étaient autorisés, être la cause de milliers d’heures de debugging à frustration garantie :)

Pour les curieux, vous pouvez en savoir plus à cet endroit de la [documentation officielle \(en anglais\)](#)

Chaînes de caractères

- Vous pouvez naturellement mettre des accents dans les chaînes de caractères. Cela dit, les données manipulées par un programme proviennent pour l’essentiel de sources externes, comme une base de données ou un formulaire web, et donc le plus souvent pas directement du code source. Les chaînes de caractères présentes dans du vrai code sont bien souvent limitées à des messages de logging, et le plus souvent d’ailleurs en anglais, donc sans accent.
- Lorsque votre programme doit interagir avec les utilisateurs et qu’il doit donc parler leur langue, c’est une bonne pratique de créer un fichier spécifique, que l’on appelle fichier de ressources, qui contient toutes les chaînes de caractères spécifiques à une langue. Ainsi, la traduction de votre programme consistera à simplement traduire ce fichier de ressources.

```
message = "on peut mettre un caractère accentué dans une chaîne"
```

Commentaires

- Enfin on peut aussi bien sûr mettre dans les commentaires n’importe quel caractère Unicode, et donc notamment des caractères accentués si on choisit malgré tout d’écrire le code en français.

```
# on peut mettre un caractère accentué dans un commentaire
# ainsi que cos(θ) ⇒ ∀x ∈ ]f(t)dt vous voyez l'idée générale
```

Qu’est-ce qu’un encodage ?

Comme vous le savez, la mémoire - ou le disque - d’un ordinateur ne permet que de stocker des représentations binaires. Il n’y a donc pas de façon “naturelle” de représenter un caractère comme ‘A’ ou un guillemet ou un point-virgule.

On utilise pour cela un encodage, par exemple le code US-ASCII - <http://www.asciitable.com/> - stipule, pour faire simple, qu’un ‘A’ est représenté par l’octet 65 qui s’écrit en binaire 01000001. Il se trouve qu’il existe plusieurs encodages, bien sûr incompatibles, selon les systèmes et les langues. Vous trouverez plus de détails ci-dessous.

Le point important est que pour pouvoir ouvrir un fichier “proprement”, il faut bien entendu disposer du **contenu** du fichier, mais il faut aussi connaître l’**encodage** qui a été utilisé pour l’écrire.

Précautions à prendre pour l’encodage de votre code source

L’encodage ne concerne pas simplement les objets chaîne de caractères, mais également votre code source. **python-3** considère que votre code source utilise **par défaut l’encodage utf-8**. Nous vous conseillons de conserver cet encodage qui est celui qui vous offrira le plus de flexibilité.

Vous pouvez malgré tout changer l’encodage **de votre code source** en faisant figurer dans vos fichiers, **en première ou deuxième ligne**, une déclaration comme ceci

```
# -*- coding: <nom_de_l_encodage> -*-
```

ou plus simplement, comme ceci

```
# coding: utf-8
```

Notons que la première option est également interprétée par l'éditeur de texte *emacs* pour utiliser le même encodage. En dehors de l'utilisation d'*emacs*, la deuxième option, plus simple et donc plus pythonique, est à préférer.

Le nom **utf-8** fait référence à **Unicode** (ou pour être précis, à l'encodage le plus répandu parmi ceux qui sont définis dans la norme Unicode, comme nous le verrons plus bas). Sur certains systèmes plus anciens vous pourrez être amenés à utiliser un autre encodage. Pour déterminer la valeur à utiliser dans votre cas précis vous pouvez faire dans l'interpréteur interactif

```
# ceci doit être exécuté sur votre machine
import sys
print(sys.getdefaultencoding())
```

Par exemple avec d'anciennes versions de Windows (en principe de plus en plus rares) vous pouvez être amenés à écrire

```
# coding: cp1252
```

La syntaxe de la ligne `coding` est précisée dans [cette documentation](#) et dans le [pep263](#).

Le grand malentendu

Si je vous envoie un fichier contenant du français encodé avec, disons, ISO-latin-15 – http://en.wikipedia.org/wiki/ISO/IEC_8859-15 – vous pouvez voir dans la table qu'un caractère '€' va être matérialisé dans mon fichier par un octet '0xA4', soit 164.

Imaginez maintenant que vous essayez d'ouvrir ce même fichier depuis un vieil ordinateur Windows configuré pour le français. Si on ne lui donne aucune indication sur l'encodage, le programme qui va lire ce fichier sur Windows va utiliser l'encodage par défaut du système, c'est-à-dire `cp1252` – <http://en.wikipedia.org/wiki/Windows-1252>. Comme vous le voyez dans cette table, l'octet '0xA4' correspond au caractère ¤ et c'est ça que vous allez voir à la place de €.

C'est à cela que sert la balise `# coding: <nom_de_l_encodage>`

De cette manière, python lorsqu'il lit votre code source, a les moyens d'interpréter correctement son contenu car il sait quel encodage utiliser. On vous rappelle que si vous ne spécifiez aucun encodage pour votre code source, python utilisera un encodage `utf-8`, ce qui est souvent le meilleur choix.

Pourquoi ça marche en local ?

Lorsque le producteur (le programme qui écrit le fichier) et le consommateur (le programme qui le lit) tournent dans le même ordinateur, tout fonctionne bien - en général - parce que les deux programmes se ramènent à l'encodage défini comme l'encodage par défaut.

On a vu pourquoi il vaut mieux toutefois être explicite, et spécifier la balise `# coding:`

Il y a une limite toutefois, si vous utilisez un linux configuré de manière minimale, il se peut qu'il utilise par défaut l'encodage `US-ASCII` - voir plus bas - qui étant très ancien ne "connaît" pas un simple é, ni a fortiori €. Pour écrire du français, il faut donc au minimum que l'encodage par défaut de votre ordinateur contienne les caractères français, comme par exemple

- `iso-latin-1`
- `iso-latin-15`
- `utf-8`

— cp1252

Un peu d'histoire sur les encodages

22.2.2 Le code US-ASCII

Jusque dans les années 1980, les ordinateurs ne parlaient pour l'essentiel que l'anglais. La première vague de standardisation avait créé l'encodage dit ASCII, ou encore US-ASCII - voir par exemple <http://www.asciitable.com>, ou en version longue <http://en.wikipedia.org/wiki/ASCII>.

Le code ASCII s'étend sur 128 valeurs, soit 7 bits, mais est le plus souvent implémenté sur un octet pour préserver l'alignement, le dernier bit pouvant être utilisé par exemple pour ajouter un code correcteur d'erreur - ce qui à l'époque des modems n'était pas superflu. Bref, la pratique courante était alors de manipuler une chaîne de caractères comme un tableau d'octets.

22.2.3 Les encodages ISO-latin

Dans les années 1990, pour satisfaire les besoins des pays européens, ont été définis plusieurs encodages alternatifs, connus sous le nom de **ISO-latin**, ou encore **ISO-8859**. Idéalement, on aurait pu et **certainement dû** définir un seul encodage pour représenter tous les nouveaux caractères, mais entre toutes les langues européennes, le nombre de caractères à ajouter était substantiel, et cet encodage unifié aurait largement dépassé 256 caractères différents, il n'aurait donc **pas été possible** de tout faire tenir sur un octet.

On a préféré préserver la “bonne propriété” du modèle *un caractère == un octet*, ceci afin de préserver le code existant qui aurait sinon dû être retouché ou réécrit.

Dès lors il n'y avait pas d'autre choix que de définir **plusieurs** encodages distincts, par exemple, pour le français on a utilisé à l'époque **ISO-latin-1**, pour le russe **ISO-latin-5**.

À ce stade, le ver était dans le fruit. Depuis cette époque pour ouvrir un fichier il faut connaître son encodage.

22.2.4 Unicode

Lorsqu'on a ensuite cherché à manipuler aussi les langues asiatiques, il a de toutes façons fallu définir de nouveaux encodages beaucoup plus larges. C'est ce qui a été fait par le standard **Unicode** qui définit 3 nouveaux encodages.

- **UTF-8** : un encodage à taille variable, à base d'octets, qui maximise la compatibilité avec ASCII,
- **UTF-16** : un encodage à taille variable, à base de mots de 16 bits
- **UTF-32** : un encodage à taille fixe, à base de mots de 32 bits

Ces 3 standards couvrent le même jeu de caractères (113 021 tout de même dans la dernière version). Parmi ceux-ci le plus utilisé est certainement `utf-8`. Un texte ne contenant que des caractères du code US-ASCII initial peut être lu avec l'encodage UTF-8.

Pour être enfin tout à fait exhaustif, si on sait qu'un fichier est au format Unicode, on peut déterminer quel est l'encodage qu'il utilise, en se basant sur les 4 premiers octets du document. Ainsi dans ce cas particulier (lorsqu'on est sûr qu'un document utilise un des trois encodages Unicode) il n'est plus nécessaire de connaître son encodage de manière “externe”.

Les outils de base sur les strings

23.1 Complément - niveau intermédiaire

23.1.1 Lire la documentation

Même après des années de pratique, il est difficile de se souvenir de toutes les méthodes travaillant sur les chaînes de caractères. Aussi il est toujours utile de recourir à la documentation embarquée

```
help(str)
```

Nous allons tenter ici de citer les méthodes les plus utilisées. Nous n'avons le temps que de les utiliser de manière très simple, mais bien souvent il est possible de passer en paramètre des options permettant de ne travailler que sur une sous-chaîne, ou sur la première ou dernière occurrence d'une sous-chaîne. Nous vous renvoyons à la documentation pour obtenir toutes les précisions utiles.

23.1.2 Découpage - assemblage : `split` et `join`

On l'a vu dans la vidéo, la paire `split` et `join` permet de découper une chaîne selon un séparateur pour obtenir une liste, et à l'inverse de reconstruire une chaîne à partir d'une liste.

`split` permet donc de découper

```
'abc:=def:=ghi:=jkl'.split('=')
```

```
['abc', 'def', 'ghi', 'jkl']
```

Et à l'inverse

```
"=" .join(['abc', 'def', 'ghi', 'jkl'])
```

```
'abc:=def:=ghi:=jkl'
```

Attention toutefois si le séparateur est un terminateur, la liste résultat contient alors une dernière chaîne vide. En pratique, on utilisera la méthode `strip`, que nous allons voir ci-dessous, avant la méthode `split` pour éviter ce problème.

```
'abc;def;ghi;jkl;'.split(';')
```

```
['abc', 'def', 'ghi', 'jkl', '']
```

Qui s'inverse correctement cependant

```
";".join(['abc', 'def', 'ghi', 'jkl', ''])
```

```
'abc;def;ghi;jkl;'
```

23.1.3 Remplacements : `replace`

`replace` est très pratique pour remplacer une sous-chaîne par une autre, avec une limite éventuelle sur le nombre de remplacements

```
"abcdefabcdefabcdef".replace("abc", "zoo")
```

```
'zoodefzoodefzoodef'
```

```
"abcdefabcdefabcdef".replace("abc", "zoo", 2)
```

```
'zoodefzoodefabcdef'
```

Plusieurs appels à `replace` peuvent être chaînés comme ceci

```
"les [x] qui disent [y]".replace("[x]", "chevaliers").replace("[y]", "Ni")
```

```
'les chevaliers qui disent Ni'
```

23.1.4 Nettoyage : `strip`

On pourrait par exemple utiliser `replace` pour enlever les espaces dans une chaîne, ce qui peut être utile pour « nettoyer » comme ceci

```
" abc:def:ghi ".replace(" ", "")
```

```
'abc:def:ghi'
```

Toutefois bien souvent on préfère utiliser `strip` qui ne s'occupe que du début et de la fin de la chaîne, et gère aussi les tabulations et autres retour à la ligne

```
" \tune chaine avec des trucs qui dépassent \n".strip()
```

```
'une chaine avec des trucs qui dépassent'
```

On peut appliquer `strip` avant `split` pour éviter le problème du dernier élément vide.

```
'abc;def;ghi;jkl;'.strip(';').split(';')
```

```
['abc', 'def', 'ghi', 'jkl']
```

23.1.5 Rechercher une sous-chaîne

Plusieurs outils permettent de chercher une sous-chaîne. Il existe `find` qui renvoie le plus petit index où on trouve la sous-chaîne

```
# l'indice du début de la première occurrence
"abcdefcdefghefghijk".find("def")
```

```
3
```

```
# ou -1 si la chaîne n'est pas présente:
"abcdefcdefghefghijk".find("zoo")
```

```
-1
```

`rfind` fonctionne comme `find` mais en partant de la fin de la chaîne

```
# en partant de la fin
"abcdefcdefghefghijk".rfind("fgh")
```

```
13
```

```
# notez que le résultat correspond
# tout de même toujours au début de la chaîne
"abcdefcdefghefghijk"[13]
```

```
'f'
```

La méthode `index` se comporte comme `find`, mais en cas d'absence elle lève une **exception** (nous verrons ce concept plus tard) plutôt que de renvoyer `-1`

```
"abcdefcdefghefghijk".index("def")
```

```
3
```

```
try:
    "abcdefcdefghefghijk".index("zoo")
except Exception as e:
    print("OOPS", type(e), e)
```

```
OOPS <class 'ValueError'> substring not found
```

Mais le plus simple pour chercher si une sous-chaîne est dans une autre chaîne est d'utiliser l'instruction `in` sur laquelle nous reviendrons lorsque nous parlerons des séquences.

```
"def" in "abcdefcdefghefghijk"
```

```
True
```

La méthode `count` compte le nombre d'occurrences d'une sous-chaîne

```
"abcdefcdefghefghijk".count("ef")
```

```
3
```

Signalons enfin les méthodes de commodité suivantes

```
"abcdefcdefghefghijk".startswith("abcd")
```

```
True
```

```
"abcdefcdefghefghijk".endswith("ghijk")
```

```
True
```

S'agissant des deux dernières, remarquons que

```
chaine.startswith(sous_chaine)  $\iff$  chaine.find(sous_chaine) == 0
```

```
chaine.endswith(sous_chaine)  $\iff$  chaine.rfind(sous_chaine) ==  
(len(chaine)-len(sous_chaine))
```

On remarque ici la supériorité en terme d'expressivité des méthodes pythoniques `startswith` et `endswith`.

23.1.6 Capitalisation

Voici pour conclure quelques méthodes utiles qui parlent d'elles-mêmes.

```
"monty PYTHON".upper()
```

```
'MONTY PYTHON'
```

```
"monty PYTHON".lower()
```

```
'monty python'
```

```
"monty PYTHON".swapcase()
```

```
'MONTY python'
```

```
"monty PYTHON".capitalize()
```

```
'Monty python'
```

```
"monty PYTHON".title()
```

```
'Monty Python'
```

23.1.7 Pour en savoir plus

Tous ces outils sont documentés en détail ici (en anglais)

Formatage de chaînes de caractères

24.1 Complément - niveau basique

On désigne par formatage les outils qui permettent d'obtenir une présentation fine des résultats, que ce soit pour améliorer la lisibilité lorsqu'on s'adresse à des humains, ou pour respecter la syntaxe d'un outil auquel on veut passer les données pour un traitement ultérieur.

24.1.1 La fonction `print`

Nous avons jusqu'à maintenant presque toujours utilisé la fonction `print` pour afficher nos résultats. Comme on l'a vu, celle-ci réalise un formatage sommaire : elle insère un espace entre les valeurs qui lui sont passées.

```
print(1, 'a', 12 + 4j)
```

```
1 a (12+4j)
```

La seule subtilité notable concernant `print` est que par défaut, elle ajoute un saut de ligne à la fin. Pour éviter ce comportement, on peut passer à la fonction un argument `end`, qui sera inséré *au lieu* du saut de ligne. Ainsi par exemple

```
# une première ligne
print("une", "seule", "ligne")
```

```
une seule ligne
```

```
# une deuxième ligne en deux appels à print
print("une", "autre", end=' ')
print("ligne")
```

```
une autre ligne
```

Il faut remarquer aussi que `print` est capable d'imprimer **n'importe quel objet**. Nous l'avons déjà fait avec les listes et les tuples, voici par exemple un module

```
# on peut imprimer par exemple un objet 'module'
import math

print('le module math est', math)
```

```
le module math est <module 'math' from '/opt/conda/lib/python3.6/lib-dynload/math.
↳cpython-36m-x86_64-linux-gnu.so'>
```

En anticipant un peu, voici comment `print` présente les instances de classe (ne vous inquiétez pas, nous apprendrons dans une semaine ultérieure ce que sont les classes et les instances).

```
# pour définir la classe Personne
class Personne:
    pass

# et pour créer une instance de cette classe
personne = Personne()
```

```
# voila comment s'affiche une instance de classe
print(personne)
```

```
<__main__.Personne object at 0x7f1b337770f0>
```

On rencontre assez vite les limites de `print`.

- D'une part, il peut être nécessaire de formater une chaîne de caractères sans nécessairement vouloir l'imprimer, ou en tous cas pas immédiatement.
- D'autre part, les espaces ajoutés peuvent être plus néfastes qu'utiles.
- Enfin, on peut avoir besoin de préciser un nombre de chiffres significatifs, ou de choisir comment présenter une date.

C'est pourquoi il est plus courant de **formater** les chaînes - c'est à dire de calculer des chaînes en mémoire, sans nécessairement les imprimer de suite, et c'est ce que nous allons étudier dans ce complément.

24.1.2 Les *f-strings*

Depuis la version 3.6 de python, on peut utiliser les *f-strings*, le premier mécanisme de formatage que nous étudierons. C'est le mécanisme de formatage le plus simple et le plus agréable à utiliser.

Je vous recommande tout de même de lire les sections à propos de `format` et de `%`, qui sont encore massivement utilisées dans le code existant (surtout `%` d'ailleurs, bien que essentiellement obsolète).

Mais définissons d'abord quelques données à afficher.

```
# donnons-nous quelques variables
prenom, nom, age = 'Jean', 'Dupont', 35
```

```
# mon premier f-string
f"{prenom} {nom} a {age} ans"
```

```
'Jean Dupont a 35 ans'
```

Vous remarquez d'abord que le string commence par `f"`, c'est bien sûr pour cela qu'on l'appelle un *f-string*.

On peut bien entendu ajouter le `f` devant toutes les formes de strings, qu'ils commencent par `'` ou `"` ou `'''` ou `"""`.

Ensuite vous remarquez que les zones délimitées entre `{ }` sont remplacées. La logique d'un *f-string*, c'est tout simplement de considérer l'intérieur d'un `{ }` comme du code python (une expression pour être précis), de l'évaluer, et d'utiliser le résultat pour remplir le `{ }`.

Ça veut dire, en clair, que je peux faire des calculs à l'intérieur des `{ }`

```
# toutes les expressions sont autorisées à l'intérieur d'un {}
f"dans 10 ans {prenom} aura {age + 10} ans"
```

```
'dans 10 ans Jean aura 45 ans'
```

```
# on peut donc aussi mettre des appels de fonction
notes = [12, 15, 19]
f"nous avons pour l'instant {len(notes)} notes"
```

```
"nous avons pour l'instant 3 notes"
```

Nous allons en rester là pour la partie en niveau basique. Il nous reste à étudier comment chaque `{ }` est formaté (par exemple comment choisir le nombre de chiffres significatifs sur un flottant), voyez plus bas pour plus de détails sur ce point.

Comme vous le voyez, les *f-strings* fournissent une méthode très simple et expressive pour formater des données dans des chaînes de caractère. Redisons-le pour être bien clair : un *f-string* **ne réalise pas d'impression**, il faut donc le passer à `print` si l'impression est souhaitée.

24.1.3 La méthode `format`

Avant l'introduction des *f-strings*, la technique recommandée pour faire du formatage était d'utiliser la méthode `format` qui est définie sur les objets `str` et qui s'utilise comme ceci

```
"{} {} a {} ans".format(prenom, nom, age)
```

```
'Jean Dupont a 35 ans'
```

Dans cet exemple le plus simple, les données sont affichées en lieu et place des `{ }`, dans l'ordre où elles sont fournies.

Cela convient bien lorsqu'on a peu de données. Si par la suite on veut changer l'ordre par exemple des nom et prénom, on peut bien sûr échanger l'ordre des arguments passés à `format`, ou encore utiliser la **liaison par position**, comme ceci

```
"{1} {0} a {2} ans".format(prenom, nom, age)
```

```
'Dupont Jean a 35 ans'
```

Dans la pratique toutefois, cette forme est assez peu utile, on lui préfère souvent la **liaison par nom** qui se présente comme ceci

```
"{le_prenom} {le_nom} a {l_age} ans".format(le_nom=nom, le_prenom=prenom, l_
↪age=age)
```

```
'Jean Dupont a 35 ans'
```

Dans ce premier exemple de liaison par nom, nous avons délibérément utilisé des noms différents pour les données externes et pour les noms apparaissant dans le format, pour bien illustrer comment la liaison est résolue, mais on peut aussi bien faire tout simplement

```
"{prenom} {nom} a {age} ans".format(nom=nom, prenom=prenom, age=age)
```

```
'Jean Dupont a 35 ans'
```

Voici qui conclut notre courte introduction à la méthode `format`.

24.2 Complément - niveau intermédiaire

24.2.1 La toute première version du formatage : l'opérateur %

`format` a été en fait introduite assez tard dans python, pour remplacer la technique que nous allons présenter maintenant.

Étant donné le volume de code qui a été écrit avec l'opérateur `%`, il nous a semblé important d'introduire brièvement cette construction ici. Vous ne devez cependant pas utiliser cet opérateur dans du code moderne, la manière pythonique de formater les chaînes de caractères est le f-string.

Le principe de l'opérateur `%` est le suivant. On élabore comme ci-dessus un « format » c'est-à-dire le patron de ce qui doit être rendu, auquel on passe des arguments pour « remplir » les trous. Voyons les exemples de tout à l'heure rendus avec l'opérateur `%`

```
# l'ancienne façon de formater les chaînes avec %  
# est souvent moins lisible  
"%s %s a %s ans" % (prenom, nom, age)
```

```
'Jean Dupont a 35 ans'
```

On pouvait également avec cet opérateur recourir à un mécanisme de liaison par nommage, en passant par un dictionnaire. Pour anticiper un tout petit peu sur cette notion que nous verrons très bientôt, voici comment

```
variables = {'le_nom' : nom, 'le_prenom' : prenom, 'l_age' : age}  
"% (le_nom)s, % (le_prenom)s, % (l_age)s ans" % variables
```

```
'Dupont, Jean, 35 ans'
```

24.3 Complément - niveau avancé

De retour aux *f-strings* et à la fonction `format`, il arrive qu'on ait besoin de spécifier plus finement la façon dont une valeur doit être affichée.

24.3.1 Précision des arrondis

C'est typiquement le cas avec les valeurs flottantes pour lesquelles la précision de l'affichage vient au détriment de la lisibilité. Voici deux formes équivalentes pour obtenir une valeur de pi arrondie :

```
from math import pi
```

```
# un f-string  
f"pi avec seulement 2 chiffres apres la virgule {pi:.2f}"
```

```
'pi avec seulement 2 chiffres apres la virgule 3.14'
```

```
# avec format avec liaison par nom  
"pi avec seulement 2 chiffres apres la virgule {flottant:.2f}".format(flottant=pi)
```

```
'pi avec seulement 2 chiffres apres la virgule 3.14'
```

Dans ces deux exemples, la partie à l'intérieur des `{ }` et à droite du `:` s'appelle le format, ici `.2f`; vous remarquerez que c'est le même pour les *f-strings* et pour `format`, et c'est toujours le cas. C'est pourquoi on ne verra plus à partir d'ici que des exemples avec les *f-strings*.

24.3.2 0 en début de nombre

Pour forcer un petit entier à s'afficher sur 5 caractères, avec des 0 ajoutés au début si nécessaire

```
x = 15

f"{x:04d}"
```

```
'0015'
```

Ici on utilise le format `d` (toutes ces lettres `d`, `f`, `g` viennent des formats ancestraux de la libc comme `printf`). Ici avec `:04d` on précise qu'on veut une sortie sur 4 caractères et qu'il faut remplir avec des 0.

24.3.3 Largeur fixe

Dans certains cas, on a besoin d'afficher des données en colonnes de largeurs fixes, on utilise pour cela les formats `< ^ et >` pour afficher à gauche, au centre, ou à droite d'une zone de largeur fixe

```
# les données à afficher
comptes = [
    ('Apollin', 'Dupont', 127),
    ('Myrtille', 'Lamartine', 25432),
    ('Prune', 'Soc', 827465),
]

for prenom, nom, solde in comptes:
    print(f"{prenom:<10} -- {nom:^12} -- {solde:>8} €")
```

```
Apollin    --      Dupont      --      127 €
Myrtille   --   Lamartine    --    25432 €
Prune      --         Soc      --   827465 €
```

24.3.4 Voir aussi

Nous vous invitons à vous reporter à la documentation de `format` pour plus de détails sur les formats disponibles, et notamment aux nombreux exemples qui y figurent.

Obtenir une réponse de l'utilisateur

25.1 Complément - niveau basique

Occasionnellement, il peut être utile de poser une question à l'utilisateur.

25.1.1 La fonction `input()`

C'est le propos de la fonction `input`. Par exemple :

```
nom_ville = input("entrez le nom de la ville : ")
print(f"nom_ville={nom_ville}")
```

```
entrez le nom de la ville : Megapolis
nom_ville=Megapolis
```

25.1.2 Attention à bien vérifier/convertir

Notez bien que `input` renvoie **toujours une chaîne**. C'est assez évident, mais il est très facile de l'oublier et de passer cette chaîne directement à une fonction qui s'attend à recevoir, par exemple, un nombre entier, auquel cas les choses se passent mal

```
>>> input("nombre de lignes ? ") + 3
nombre de lignes ? 12
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: must be str, not int
```

Dans ce cas il faut appeler la fonction `int` pour convertir le résultat en un entier

```
int(input("nombre de lignes ? ")) + 3
```

```
nombre de lignes ? 42
```

```
45
```

25.1.3 Limitations

Cette fonction peut être utile pour vos premiers pas en python.

En pratique toutefois, on utilise assez peu cette fonction, car les applications « réelles » viennent avec leur propre interface utilisateur, souvent graphique, et disposent donc d'autres moyens que celui-ci pour interagir avec l'utilisateur.

Les applications destinées à fonctionner dans un terminal, quant à elles, reçoivent traditionnellement leurs données de la ligne de commande. C'est le propos du module `argparse` que nous avons déjà rencontré en 1^{re} semaine.

Expressions régulières et le module `re`

26.1 Complément - niveau intermédiaire

Une expression régulière est un objet mathématique permettant de décrire un ensemble de textes qui possèdent des propriétés communes. Par exemple, s'il vous arrive d'utiliser un terminal, et que vous tapez

```
$ dir *.txt
```

(ou `ls *.txt` sur linux ou mac), vous utilisez l'expression régulière `*.txt` qui désigne tous les fichiers dont le nom se termine par `.txt`. On dit que l'expression régulière *filtre* toutes les chaînes qui se terminent par `.txt` (l'expression anglaise consacrée est le *pattern matching*).

Le langage Perl a été le premier à populariser l'utilisation des expressions régulières en les supportant nativement dans le langage, et non au travers d'une librairie.

En python, les expressions régulières sont disponibles de manière plus traditionnelle, via le module `re` de la librairie standard, que nous allons voir maintenant.

Dans la commande ci-dessus, `*.txt` est une expression régulière très simple. Le module `re` fournit le moyen de construire des expressions régulières très élaborées et plus puissantes que ce que supporte le terminal. C'est pourquoi la syntaxe des regexps de `re` est un peu différente. Par exemple, pour filtrer la même famille de chaînes que `*.txt` avec le module `re`, il nous faudra écrire l'expression régulière sous une forme légèrement différente.

Le propos de ce complément est de vous donner une première introduction au module `re`.

```
import re
```

Je vous conseille d'avoir sous la main la *documentation du module 're'* <<https://docs.python.org/3/library/re.html>> pendant que vous lisez ce complément.

26.1.1 Avertissement

Dans ce complément nous serons amenés à utiliser des traits qui dépendent du LOCALE, c'est-à-dire, pour faire simple, de la configuration de l'ordinateur vis-à-vis de la langue.

Tant que vous exécutez ceci dans le notebook sur la plateforme, en principe tout le monde verra exactement la même chose. Par contre, si vous faites tourner le même code sur votre ordinateur, il se peut que vous obteniez des résultats légèrement différents.

26.1.2 Un exemple simple

findall

On se donne deux exemples de chaînes

```
sentences = ['Lacus a donec, vitae gravida proin sociis.',
             'Neque ipsum! rhoncus cras quam.']
```

On peut **chercher tous** les mots se terminant par a ou m dans une chaîne avec `findall`

```
for sentence in sentences:
    print(f"---- dans >{sentence}<")
    print(re.findall(r"\w*[am]\W", sentence))
```

```
---- dans >Lacus a donec, vitae gravida proin sociis.<
['a ', 'gravida ']
---- dans >Neque ipsum! rhoncus cras quam.<
['ipsum!', 'quam.']
```

Ce code permet de chercher toutes (`findall`) les occurrences de l'expression régulière, qui ici est définie par le *raw-string*

```
r"\w*[am]\W"
```

Nous verrons tout à l'heure comment fabriquer des expressions régulières plus en détail, mais pour démystifier au moins celle-ci, on a mis bout à bout les morceaux suivants.

- `\w*` : on veut trouver une sous-chaîne qui commence par un nombre quelconque, y compris nul (*) de caractères alphanumériques (`\w`). Ceci est défini en fonction de votre LOCALE, on y reviendra.
- `[am]` : immédiatement après, il nous faut trouver un caractère a ou m.
- `\W` : et enfin, il nous faut un caractère qui ne soit **pas** alphanumérique. Ceci est important puisqu'on cherche les mots qui **se terminent** par un a ou un m, si on ne le mettait pas on obtiendrait ceci

```
# le \W final est important
# voici ce qu'on obtient si on l'omet
for sentence in sentences:
    print(f"---- dans >{sentence}<")
    print(re.findall(r"\w*[am]", sentence))
```

```
---- dans >Lacus a donec, vitae gravida proin sociis.<
['La', 'a', 'vita', 'gravida']
---- dans >Neque ipsum! rhoncus cras quam.<
['ipsum', 'cra', 'quam']
```

split

Une autre forme simple d'utilisation des regexps est `re.split`, qui fournit une fonctionnalité voisine de `str.split`, mais où les séparateurs sont exprimés comme une expression régulière

```
for sentence in sentences:
    print(f"---- dans >{sentence}<")
    print(re.split(r"\W+", sentence))
    print()
```



```

---- dans >Lacus a donec, vitae gravida proin sociis.<
['Lacus', 'a', 'donec', 'vitae', 'gravida', 'proin', 'sociis', '']

---- dans >Neque ipsum! rhoncus cras quam.<
['Neque', 'ipsum', 'rhoncus', 'cras', 'quam', '']

```

Ici l'expression régulière, qui bien sûr décrit le séparateur, est simplement `\W+` c'est-à-dire toute suite d'au moins un caractère non alphanumérique.

Nous avons donc là un moyen simple, et plus puissant que `str.split`, de couper un texte en mots.

sub

Une troisième méthode utilitaire est `re.sub` qui permet de remplacer les occurrences d'une *regex*, comme par exemple

```

for sentence in sentences:
    print(f"---- dans >{sentence}<")
    print(re.sub(r"(\w+)", r"X\1Y", sentence))
    print()

```

```

---- dans >Lacus a donec, vitae gravida proin sociis.<
XLacusY XaY XdonecY, XvitaeY XgravidaY XproinY XsociisY.

---- dans >Neque ipsum! rhoncus cras quam.<
XNequeY XipsumY! XrhoncusY XcrasY XquamY.

```

Ici, l'expression régulière (le premier argument) contient un **groupe** : on a utilisé des parenthèses autour du `\w+`. Le second argument est la chaîne de remplacement, dans laquelle on a fait **référence au groupe** en écrivant `\1`, qui veut dire tout simplement « le premier groupe ».

Donc au final, l'effet de cet appel est d'entourer toutes les suites de caractères alphanumériques par X et Y.

Pourquoi un *raw-string* ?

En guise de digression, il n'y a aucune obligation à utiliser un *raw-string*, d'ailleurs on rappelle qu'il n'y a pas de différence de nature entre un *raw-string* et une chaîne usuelle

```

raw = r'abc'
regular = 'abc'
# comme on a pris une 'petite' chaîne ce sont les mêmes objets
print(f"both compared with is → {raw is regular}")
# et donc a fortiori
print(f"both compared with == → {raw == regular}")

```

```

both compared with is → True
both compared with == → True

```

Il se trouve que le *backslash* `\` à l'intérieur des expressions régulières est d'un usage assez courant - on l'a vu déjà plusieurs fois. C'est pourquoi on **utilise fréquemment un *raw-string*** pour décrire une expression régulière, et en général à chaque fois qu'elle comporte une *backslash*. On rappelle que le *raw-string* désactive l'interprétation des `\` à l'intérieur de la chaîne, par exemple, `\t` est interprété comme un caractère de tabulation. Sans *raw-string*, il faut doubler tous les `\` pour qu'il n'y ait pas d'interprétation.

26.1.3 Un deuxième exemple

Nous allons maintenant voir comment on peut d'abord vérifier si une chaîne est conforme au critère défini par l'expression régulière, mais aussi *extraire* les morceaux de la chaîne qui correspondent aux différentes parties de

l'expression.

Pour cela, supposons qu'on s'intéresse aux chaînes qui comportent 5 parties, une suite de chiffres, une suite de lettres, des chiffres à nouveau, des lettres et enfin de nouveau des chiffres.

Pour cela on considère ces trois chaînes en entrée

```
inputs = ['890hj000nnm890',    # cette entrée convient
          '123abc456def789',    # celle-ci aussi
          '8090abababab879',    # celle-ci non
          ]
```

match

Pour commencer, voyons que l'on peut facilement **vérifier si une chaîne vérifie** ou non le critère.

```
regexpl = "[0-9]+[A-Za-z]+[0-9]+[A-Za-z]+[0-9]+"
```

Si on applique cette expression régulière à toutes nos entrées

```
for input in inputs:
    match = re.match(regexpl, input)
    print(f"{input:16s} → {match}")
```

```
890hj000nnm890    → <_sre.SRE_Match object; span=(0, 14), match='890hj000nnm890'>
123abc456def789    → <_sre.SRE_Match object; span=(0, 15), match='123abc456def789'>
8090abababab879    → None
```

Pour rendre ce résultat un peu plus lisible nous nous définissons une petite fonction de confort.

```
# pour simplement visualiser si on a un match ou pas
def nice(match):
    # le retour de re.match est soit None, soit un objet match
    return "no" if match is None else "Match!"
```

Avec quoi on peut refaire l'essai sur toutes nos entrées.

```
# la même chose mais un peu moins encombrant
print(f"REGEXP={regexpl}\n")
for input in inputs:
    match = re.match(regexpl, input)
    print(f"{input:>16s} → {nice(match)}")
```

```
REGEXP=[0-9]+[A-Za-z]+[0-9]+[A-Za-z]+[0-9]+
```

```
890hj000nnm890 → Match!
123abc456def789 → Match!
8090abababab879 → no
```

Ici plutôt que d'utiliser les raccourcis comme `\w` j'ai préféré écrire explicitement les ensembles de caractères en jeu. De cette façon, on rend son code indépendant du LOCALE si c'est ce qu'on veut faire. Il y a deux morceaux qui interviennent tour à tour :

- `[0-9]+` signifie une suite de au moins un caractère dans l'intervalle `[0-9]`,
- `[A-Za-z]+` pour une suite d'au moins un caractère dans l'intervalle `[A-Z]` ou dans l'intervalle `[a-z]`.

Et comme tout à l'heure on a simplement juxtaposé les morceaux dans le bon ordre pour construire l'expression régulière complète.

Nommer un morceau (un groupe)

```
# on se concentre sur une entrée correcte
haystack = inputs[1]
haystack
```

```
'123abc456def789'
```

Maintenant, on va même pouvoir **donner un nom** à un morceau de la regexp, ici on désigne par `needle` le groupe de chiffres du milieu.

```
# la même regexp, mais on donne un nom au groupe de chiffres central
regexp2 = "[0-9]+[A-Za-z]+(?P<needle>[0-9]+)[A-Za-z]+[0-9]+"
```

Et une fois que c'est fait, on peut demander à l'outil de nous **retrouver la partie correspondante** dans la chaîne initiale :

```
print(re.match(regexp2, haystack).group('needle'))
```

```
456
```

Dans cette expression on a utilisé un **groupe nommé** `(?P<needle>[0-9]+)`, dans lequel :

- les parenthèses définissent un groupe,
- `?P<needle>` spécifie que ce groupe pourra être référencé sous le nom `needle` (cette syntaxe très absconse est héritée semble-t-il de perl).

26.1.4 Un troisième exemple

Enfin, et c'est un trait qui n'est pas présent dans tous les langages, on peut restreindre un morceau de chaîne à être identique à un groupe déjà vu plus tôt dans la chaîne. Dans l'exemple ci-dessus, on pourrait ajouter comme contrainte que le premier et le dernier groupes de chiffres soient identiques, comme ceci

```
regexp3 = "(?P<id>[0-9]+)[A-Za-z]+(?P<needle>[0-9]+)[A-Za-z]+(?P=id)"
```

Si bien que maintenant, avec les mêmes entrées que tout à l'heure

```
print(f"REGEXP={regexp3}\n")
for input in inputs:
    match = re.match(regexp3, input)
    print(f"{input:>16s} → {nice(match)}")
```

```
REGEXP=(?P<id>[0-9]+)[A-Za-z]+(?P<needle>[0-9]+)[A-Za-z]+(?P=id)
```

```
890hj000nnm890 → Match!
123abc456def789 → no
8090abababab879 → no
```

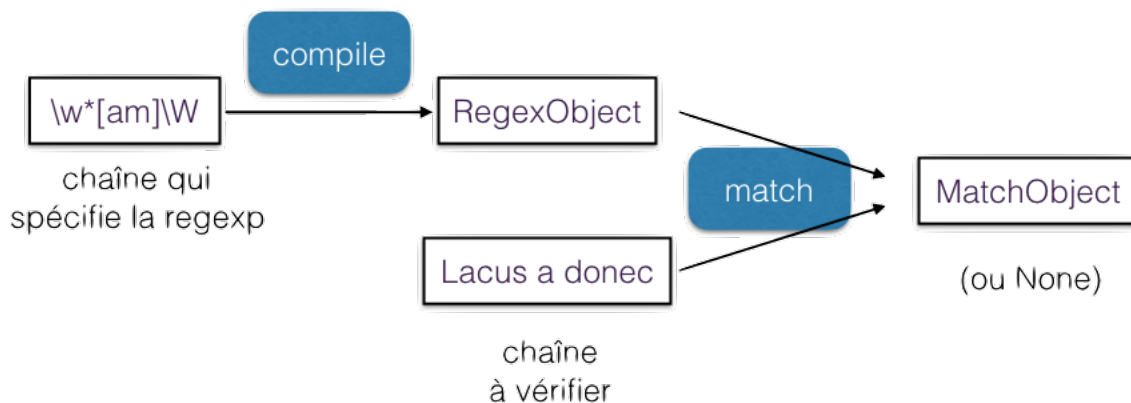
Comme précédemment on a défini le groupe nommé `id` comme étant la première suite de chiffres. La nouveauté ici est la **contrainte** qu'on a imposée sur le dernier groupe avec `(?P=id)`. Comme vous le voyez, on n'obtient un *match* qu'avec les entrées dans lesquelles le dernier groupe de chiffres est identique au premier.

26.1.5 Comment utiliser la librairie

Avant d'apprendre à écrire une expression régulière, disons quelques mots du mode d'emploi de la librairie.

Fonctions de commodité et *workflow*

Comme vous le savez peut-être, une expression régulière décrite sous forme de chaîne, comme par exemple `"\w*[am]\W"`, peut être traduite dans un **automate fini** qui permet de faire le filtrage avec une chaîne. C'est ce qui explique le *workflow* que nous avons résumé dans cette figure.



La méthode recommandée pour utiliser la librairie, lorsque vous avez le même *pattern* à appliquer à un grand nombre de chaînes, est de :

- compiler **une seule fois** votre chaîne en un automate, qui est matérialisé par un objet de la classe `re.RegexObject`, en utilisant `re.compile`,
- puis d'**utiliser directement cet objet** autant de fois que vous avez de chaînes.

Nous avons utilisé dans les exemples plus haut (et nous continuerons plus bas pour une meilleure lisibilité) des **fonctions de commodité** du module, qui sont pratiques, par exemple, pour mettre au point une expression régulière en mode interactif, mais qui ne **sont pas forcément** adaptées dans tous les cas.

Ces fonctions de commodité fonctionnent toutes sur le même principe :

```
re.match(regex, input)  $\iff$  re.compile(regex).match(input)
```

Donc à chaque fois qu'on utilise une fonction de commodité, on recompile la chaîne en automate, ce qui, dès qu'on a plus d'une chaîne à traiter, représente un surcoût.

```
# au lieu de faire comme ci-dessus:

# imaginez 10**6 chaînes dans inputs
for input in inputs:
    match = re.match(regex3, input)
    print(f"{input:>16s} → {nice(match)}")
```

```
890hj000nm890 → Match!
123abc456def789 → no
8090abababab879 → no
```

```
# dans du vrai code on fera plutôt:

# on compile la chaîne en automate une seule fois
re_obj3 = re.compile(regex3)

# ensuite on part directement de l'automate
for input in inputs:
    match = re_obj3.match(input)
    print(f"{input:>16s} → {nice(match)}")
```

```
890hj000nnm890 → Match!
123abc456def789 → no
8090abababab879 → no
```

Cette deuxième version ne compile qu’une fois la chaîne en automate, et donc est plus efficace.

Les méthodes sur la classe `RegexObject`

Les objets de la classe `RegexObject` représentent donc l’automate à état fini qui est le résultat de la compilation de l’expression régulière. Pour résumer ce qu’on a déjà vu, les méthodes les plus utiles sur un objet `RegexObject` sont :

- `match` et `search`, qui cherchent une *match* soit uniquement au début (`match`) ou n’importe où dans la chaîne (`search`),
- `findall` et `split` pour chercher toutes les occurrences (`findall`) ou leur négatif (`split`),
- `sub` (qui aurait pu sans doute s’appeler `replace`, mais c’est comme ça) pour remplacer les occurrences de `pattern`.

Exploiter le résultat

Les **méthodes** disponibles sur la classe “`re.MatchObject`” sont [documentées en détail ici](#). On en a déjà rencontré quelques-unes, en voici à nouveau un aperçu rapide.

```
# exemple
input = "    Isaac Newton, physicist"
match = re.search(r"(\w+) (?P<name>\w+)", input)
```

`re` et `string` pour retrouver les données d’entrée du `match`.

```
match.string
```

```
'    Isaac Newton, physicist'
```

```
match.re
```

```
re.compile(r'(\w+) (?P<name>\w+)', re.UNICODE)
```

`group`, `groups`, `groupdict` pour retrouver les morceaux de la chaîne d’entrée qui correspondent aux **groupes** de la regex. On peut y accéder par rang, ou par nom (comme on l’a vu plus haut avec `needle`).

```
match.groups()
```

```
('Isaac', 'Newton')
```

```
match.group(1)
```

```
'Isaac'
```

```
match.group('name')
```

```
'Newton'
```

```
match.group(2)
```

```
'Newton'
```

```
match.groupdict()
```

```
{'name': 'Newton'}
```

Comme on le voit pour l'accès par rang **les indices commencent à 1** pour des raisons historiques (on peut déjà référencer \1 en sed depuis la fin des années 70).

On peut aussi accéder au **groupe 0** comme étant la partie de la chaîne de départ qui a effectivement été filtrée par l'expression régulière, et qui peut tout à fait être au beau milieu de la chaîne de départ, comme dans notre exemple

```
match.group(0)
```

```
'Isaac Newton'
```

expand permet de faire une espèce de `str.format` avec les valeurs des groupes.

```
match.expand(r"last_name \g<name> first_name \1")
```

```
'last_name Newton first_name Isaac'
```

span pour connaître les index dans la chaîne d'entrée pour un groupe donné.

```
begin, end = match.span('name')
input[begin:end]
```

```
'Newton'
```

Les différents modes (*flags*)

Enfin il faut noter qu'on peut passer à `re.compile` un certain nombre de *flags* qui modifient globalement l'interprétation de la chaîne, et qui peuvent rendre service.

Vous trouverez [une liste exhaustive de ces *flags* ici](#). Ils ont en général un nom long et parlant, et un alias court sur un seul caractère. Les plus utiles sont sans doute :

- IGNORECASE (*alias* I) pour, eh bien, ne pas faire la différence entre minuscules et majuscules,
- UNICODE (*alias* U) pour rendre les séquences \w et autres basées sur les propriétés des caractères dans la norme Unicode,
- LOCALE (*alias* L) cette fois \w dépend du locale courant,
- MULTILINE (*alias* M), et
- DOTALL (*alias* S) pour ces deux flags voir la discussion à la fin du complément.

Comme c'est souvent le cas, on doit passer à `re.compile` un **ou logique** (caractère |) des différents flags que l'on veut utiliser, c'est-à-dire qu'on fera par exemple

```
regexp = "a*b+"
re_obj = re.compile(regexp, flags=re.IGNORECASE | re.DEBUG)
```

```
MAX_REPEAT 0 MAXREPEAT
LITERAL 97
MAX_REPEAT 1 MAXREPEAT
LITERAL 98
```

```
# on ignore la casse des caractères
print(regex, "->", nice(re_obj.match("AabB")))
```

```
a*b+ -> Match!
```

26.1.6 Comment construire une expression régulière

Nous pouvons à présent voir comment construire une expression régulière, en essayant de rester synthétique (la documentation du module ‘re’ <<https://docs.python.org/3/library/re.html>> en donne une version exhaustive).

La brique de base : le caractère

Au commencement il faut spécifier des caractères.

- **un seul** caractère :
 - vous le citez tel quel, en le précédant d’un backslash \ s’il a par ailleurs un sens spécial dans le micro-langage de regexps (comme +, *, [, etc.);
- l’**attrape-tout** (*wildcard*) :
 - un point . signifie « n’importe quel caractère »;
- **un ensemble** de caractères avec la notation [. . .] qui permet de décrire par exemple :
 - [a1=] un ensemble in extenso, ici un caractère parmi a, 1, ou =,
 - [a-z] un intervalle de caractères, ici de a à z,
 - [15e-g] un mélange des deux, ici un ensemble qui contiendrait 1, 5, e, f et g,
 - [^15e-g] une **négation**, qui a ^ comme premier caractère dans les [], ici tout sauf l’ensemble précédent;
- un **ensemble prédéfini** de caractères, qui peuvent alors dépendre de l’environnement (UNICODE et LOCALE) avec entre autres les notations :
 - \w les caractères alphanumériques, et \W (les autres),
 - \s les caractères « blancs » - espace, tabulation, saut de ligne, etc., et \S (les autres),
 - \d pour les chiffres, et \D (les autres).

```
input = "abcd"

for regexp in ['abcd', 'ab[cd][cd]', 'ab[a-z]d', r'abc.', r'abc\.']:
    match = re.match(regexp, input)
    print(f"{input} / {regexp:<10s} → {nice(match)}")
```

```
abcd / abcd          → Match!
abcd / ab[cd][cd]    → Match!
abcd / ab[a-z]d      → Match!
abcd / abc.          → Match!
abcd / abc\.         → no
```

Pour ce dernier exemple, comme on a backslashé le . il faut que la chaîne en entrée contienne vraiment un .

```
print(nice(re.match(r"abc\.", "abc.")))
```

```
Match!
```

En série ou en parallèle

Si je fais une analogie avec les montages électriques, jusqu’ici on a vu le montage en série, on met des expressions régulières bout à bout qui filtrent (*match*) la chaîne en entrée séquentiellement du début à la fin. On a *un peu* de marge pour spécifier des alternatives, lorsqu’on fait par exemple

```
"ab[cd]ef"
```

mais c'est limité à **un seul** caractère. Si on veut reconnaître deux mots qui n'ont pas grand-chose à voir comme abc **ou** def, il faut en quelque sorte mettre deux regexps en parallèle, et c'est ce que permet l'opérateur |

```
regexp = "abc|def"

for input in ['abc', 'def', 'aef']:
    match = re.match(regexp, input)
    print(f"{input} / {regexp} → {nice(match)}")
```

```
abc / abc|def → Match!
def / abc|def → Match!
aef / abc|def → no
```

Fin(s) de chaîne

Selon que vous utilisez `match` ou `search`, vous précisez si vous vous intéressez uniquement à un match en début (`match`) ou n'importe où (`search`) dans la chaîne.

Mais indépendamment de cela, il peut être intéressant de « coller » l'expression en début ou en fin de ligne, et pour ça il existe des caractères spéciaux :

- `^` lorsqu'il est utilisé comme un caractère (c'est à dire pas en début de `[]`) signifie un début de chaîne ;
- `\A` a le même sens (sauf en mode MULTILINE), et je le recommande de préférence à `^` qui est déjà pas mal surchargé ;
- `$` matche une fin de ligne ;
- `\Z` est voisin mais pas tout à fait identique.

Reportez-vous à la documentation pour le détails des différences. Attention aussi à entrer le `^` correctement, il vous faut le caractère ASCII et non un voisin dans la ménagerie Unicode.

```
input = 'abcd'

for regexp in [ 'bc', r'\Aabc', '^abc', r'\Abc', '^bc', r'bcd\Z', 'bcd$', r'bc\Z',
    ↪ 'bc$' ]:
    match = re.match(regexp, input)
    print(f"{input} / {regexp:5s} → {nice(match)}")
```

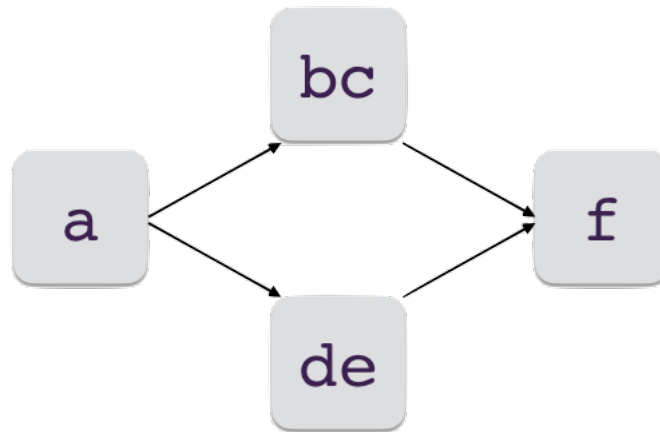
```
abcd / bc      → no
abcd / Aabc    → Match!
abcd / ^abc    → Match!
abcd / Abc     → no
abcd / ^bc     → no
abcd / bcdZ    → no
abcd / bcd$    → no
abcd / bcZ     → no
abcd / bc$     → no
```

On a en effet bien le pattern `bc` dans la chaîne en entrée, mais il n'est ni au début ni à la fin.

Parenthéser - (grouper)

Pour pouvoir faire des montages élaborés, il faut pouvoir parenthéser.


```
# une parenthèse dans une RE
# pour mettre en ligne:
# un début 'a',
# un milieu 'bc' ou 'de'
# et une fin 'f'
regex = "a(bc|de)f"
```



```
for input in ['abcf', 'adef', 'abef', 'abf']:
    match = re.match(regex, input)
    print(f"{input:>4s} → {nice(match)}")
```

```
abcf → Match!
adef → Match!
abef → no
abf  → no
```

Les parenthèses jouent un rôle additionnel de **groupe**, ce qui signifie qu'on **peut retrouver** le texte correspondant à l'expression régulière comprise dans les `()`. Par exemple, pour le premier match

```
input = 'abcf'
match = re.match(regex, input)
print(f"{input}, {regex} → {match.groups()}")
```

```
abcf, a(bc|de)f → ('bc',)
```

dans cet exemple, on n'a utilisé qu'un seul groupe `()`, et le morceau de chaîne qui correspond à ce groupe se trouve donc être le seul groupe retourné par `MatchObject.group`.

Compter les répétitions

Vous disposez des opérateurs suivants :

- `*` l'étoile qui signifie n'importe quel nombre, même nul, d'occurrences - par exemple, `(ab)*` pour indiquer '' ou 'ab' ou 'abab' ou etc.,
- `+` le plus qui signifie au moins une occurrence - e.g. `(ab)+` pour ab ou abab ou ababab ou etc,
- `?` qui indique une option, c'est-à-dire 0 ou 1 occurrence - autrement dit `(ab)?` matche '' ou ab,
- `{n}` pour exactement n occurrences de `(ab)` - e.g. `(ab){3}` qui serait exactement équivalent à ababab,
- `{m,n}` entre m et n fois inclusivement.

```
inputs = [n*'ab' for n in [0, 1, 3, 4]] + ['baba']

for regex in ['(ab)*', '(ab)+', '(ab){3}', '(ab){3,4}']:
```

```
# on ajoute \A \Z pour matcher toute la chaine
line_regexp = r"\A{}\Z".format(regex)
for input in inputs:
    match = re.match(line_regexp, input)
    print(f"{input:>8s} / {line_regexp:14s} → {nice(match)}")
```

```
      / A(ab)*Z      → Match!
ab / A(ab)*Z      → Match!
ababab / A(ab)*Z   → Match!
abababab / A(ab)*Z → Match!
baba / A(ab)*Z     → no
      / A(ab)+Z     → no
ab / A(ab)+Z       → Match!
ababab / A(ab)+Z   → Match!
abababab / A(ab)+Z → Match!
baba / A(ab)+Z     → no
      / A(ab){3}Z   → no
ab / A(ab){3}Z     → no
ababab / A(ab){3}Z → Match!
abababab / A(ab){3}Z → no
baba / A(ab){3}Z   → no
      / A(ab){3,4}Z → no
ab / A(ab){3,4}Z   → no
ababab / A(ab){3,4}Z → Match!
abababab / A(ab){3,4}Z → Match!
baba / A(ab){3,4}Z → no
```

Groupes et contraintes

Nous avons déjà vu un exemple de groupe nommé (voir `needle` plus haut), les opérateurs que l'on peut citer dans cette catégorie sont :

- `(...)` les parenthèses définissent un groupe anonyme,
- `(?P<name>...)` définit un groupe nommé,
- `(?:...)` permet de mettre des parenthèses mais sans créer un groupe, pour optimiser l'exécution puisqu'on n'a pas besoin de conserver les liens vers la chaîne d'entrée,
- `(?P=name)` qui ne matche que si l'on retrouve à cet endroit de l'entrée la même sous-chaîne que celle trouvée pour le groupe `name` en amont,
- enfin `(?=...)`, `(?!...)` et `(?<=...)` permettent des contraintes encore plus élaborées, nous vous laissons le soin d'expérimenter avec elles si vous êtes intéressés ; sachez toutefois que l'utilisation de telles constructions peut en théorie rendre l'interprétation de votre expression régulière beaucoup moins efficace.

Greedy vs non-greedy

Lorsqu'on stipule une répétition un nombre indéfini de fois, il se peut qu'il existe **plusieurs** façons de filtrer l'entrée avec l'expression régulière. Que ce soit avec `*`, ou `+`, ou `?`, l'algorithme va toujours essayer de trouver la **séquence la plus longue**, c'est pourquoi on qualifie l'approche de *greedy* - quelque chose comme glouton en français.

```
# un fragment d'HTML
line='<h1>Title</h1>'

# si on cherche un texte quelconque entre crochets
# c'est-à-dire l'expression régulière "<.*>"
re_greedy = '<.*>'

# on obtient ceci
```

```
# on rappelle que group(0) montre la partie du fragment
# HTML qui matche l'expression régulière
match = re.match(re_greedy, line)
match.group(0)
```

```
'<h1>Title</h1>'
```

Ça n'est pas forcément ce qu'on voulait faire, aussi on peut spécifier l'approche inverse, c'est-à-dire de trouver la **plus-petite** chaîne qui matche, dans une approche dite *non-greedy*, avec les opérateurs suivants :

- `*? : *` mais *non-greedy*,
- `+? : +` mais *non-greedy*,
- `?? : ?` mais *non-greedy*,

```
# ici on va remplacer * par *? pour rendre l'opérateur * non-greedy
re_non_greedy = re_greedy = '<.*?>'

# mais on continue à chercher un texte entre <> naturellement
# si bien que cette fois, on obtient
match = re.match(re_non_greedy, line)
match.group(0)
```

```
'<h1>'
```

S'agissant du traitement des fins de ligne

Il peut être utile, pour conclure cette présentation, de préciser un peu le comportement de la librairie vis-à-vis des fins de ligne.

Historiquement, les expressions régulières telles qu'on les trouve dans les librairies C, donc dans `sed`, `grep` et autres utilitaires Unix, sont associées au modèle mental où on filtre les entrées ligne par ligne.

Le module `re` en garde des traces, puisque

```
# un exemple de traitement des 'newline'
input = """une entrée
sur
plusieurs
lignes
"""
```

```
match = re.compile("(.*)").match(input)
match.groups()
```

```
('une entrée',)
```

Vous voyez donc que l'attrape-tout `'.'` en fait n'attrape pas le caractère de fin de ligne `\n`, puisque si c'était le cas et compte tenu du côté *greedy* de l'algorithme on devrait voir ici tout le contenu de `input`. Il existe un *flag* `re.DOTALL` qui permet de faire de `.` un vrai attrape-tout qui capture aussi les *newline*

```
match = re.compile("(.*)", flags=re.DOTALL).match(input)
match.groups()
```

```
('une entréensurnplusieursnlinesn',)
```

Cela dit, le caractère *newline* est par ailleurs considéré comme un caractère comme un autre, on peut le mentionner **dans une regexp** comme les autres. Voici quelques exemples pour illustrer tout ceci

```
# sans mettre le flag unicode \w ne matche que l'ASCII
match = re.compile("([\w ]*)").match(input)
match.groups()
```

```
('une entrée',)
```

```
# sans mettre le flag unicode \w ne matche que l'ASCII
match = re.compile("([\w ]*)", flags=re.U).match(input)
match.groups()
```

```
('une entrée',)
```

```
# si on ajoute \n à la liste des caractères attendus
# on obtient bien tout le contenu initial

# attention ici il ne FAUT PAS utiliser un raw string,
# car on veut vraiment écrire un newline dans la regexp

match = re.compile("([\w \n]*)", flags=re.UNICODE).match(input)
match.groups()
```

```
('une entréensurnplusieursnlinesn',)
```

26.1.7 Conclusion

La mise au point d'expressions régulières est certes un peu exigeante, et demande pas mal de pratique, mais permet d'écrire en quelques lignes des fonctionnalités très puissantes, c'est un investissement très rentable :)

Je vous signale enfin l'existence de **sites web** qui évaluent une expression régulière **de manière interactive** et qui peuvent rendre la mise au point moins fastidieuse.

Je vous signale notamment <https://pythex.org/>, et il en existe beaucoup d'autres.

26.1.8 Pour en savoir plus

Pour ceux qui ont quelques rudiments de la théorie des langages, vous savez qu'on distingue en général

- l'**analyse lexicale**, qui découpe le texte en morceaux (qu'on appelle des *tokens*),
- et l'**analyse syntaxique** qui décrit pour simplifier à l'extrême l'ordre dans lequel on peut trouver les tokens.

Avec les expressions régulières, on adresse le niveau de l'analyse lexicale. Pour l'analyse syntaxique, qui est franchement au delà des objectifs de ce cours, il existe de nombreuses alternatives, parmi lesquelles :

- `pyparsing` <<http://pyparsing.wikispaces.com/Download+and+Installation>>'
- `PLY` (Python Lex-Yacc) <<http://www.dabeaz.com/ply/>>'
- `ANTLR` <<http://www.antlr.org>>' qui est un outil écrit en Java mais qui peut générer des parsers en python,
- ...

CHAPITRE 27

Expressions régulières

Nous vous proposons dans ce notebook quelques exercices sur les expressions régulières. Faisons quelques remarques avant de commencer :

- nous nous concentrons sur l'écriture de l'expression régulière en elle-même, et pas sur l'utilisation de la librairie ;
- en particulier, tous les exercices font appel à `re.match` entre votre *regex* et une liste de chaînes d'entrée qui servent de jeux de test.

27.1 Liens utiles

Pour travailler sur ces exercices, vous pouvez profitablement avoir sous la main :

- la [documentation officielle](#),
- et [cet outil interactif sur https://pythex.org/](https://pythex.org/) qui permet d'avoir un retour presque immédiat, et donc d'accélérer la mise au point.

27.1.1 Exercice - niveau basique

27.2 identificateurs python

```
# évaluez cette cellule pour charger l'exercice
from corrections.regexpythonid import exo_pythonid
```

Nous avons vu au début de la semaine 2 que les variables en python commencent par une lettre ou un `_`, suivi de lettres, chiffres ou `_`.

Écrivez une expression régulière qui décrit ce qu'on peut utiliser comme nom de variable en python.

```
# quelques exemples de résultat attendus
exo_pythonid.example()
```

```
# à vous de jouer: écrivez ici
# sous forme de chaîne votre expression régulière

regex_pythonid = r"<votre_regex>"
```

```
# évaluez cette cellule pour valider votre code
exo_pythonid.correction(regex_pythonid)
```

27.2.1 Exercice - niveau intermédiaire (1)

27.3 lignes avec nom et prénom

```
# pour charger l'exercice
from corrections.regex_agenda import exo_agenda
```

On veut reconnaître dans un fichier toutes les lignes qui contiennent un nom et un prénom.

```
exo_agenda.example()
```

Plus précisément, on cherche les chaînes qui

- commencent par une suite - possiblement vide - de caractères alphanumériques (vous pouvez utiliser `\w`) ou tiret haut (`-`) qui constitue le prénom,
- contiennent ensuite comme séparateur le caractère “deux-points” :
- contiennent ensuite une suite - cette fois jamais vide - de caractères alphanumériques, qui constitue le nom,
- et enfin contiennent un deuxième caractère : mais optionnellement seulement.

On vous demande de construire une expression régulière qui définit les deux groupes nom et prénom, et qui rejette les lignes qui ne satisfont pas ces critères.

```
# entrez votre regex ici
# il faudra la faire terminer par \Z
# regardez ce qui se passe si vous ne le faites pas

regex_agenda = r"<votre_regex>\Z"
```

```
# évaluez cette cellule pour valider votre code
exo_agenda.correction(regex_agenda)
```

27.3.1 Exercice - niveau intermédiaire (2)

27.4 numéros de téléphone

```
# pour charger l'exercice
from corrections.regex_phone import exo_phone
```

Cette fois on veut reconnaître des numéros de téléphone français, qui peuvent être :

- soit au format contenant 10 chiffres dont le premier est un 0,
- soit un format international commençant par +33 suivie de 9 chiffres.

Dans tous les cas on veut trouver dans le groupe “number” les 9 chiffres vraiment significatifs, comme ceci :

```
exo_phone.example()
```

```
# votre regexp
# à nouveau il faut terminer la regexp par \Z
regexp_phone = r"<votre regexp>\Z"
```

```
# évaluez cette cellule pour valider votre code
exo_phone.correction(regexp_phone)
```

27.4.1 Exercice - niveau avancé

Vu comment sont conçus les exercices, vous ne pouvez pas passer à `re.compile` un *flag* comme `re.IGNORECASE` ou autre ; sachez cependant que vous pouvez ***embarquer* ces flags dans la *regexp*** elle-même ; par exemple pour rendre la regexp insensible à la casse de caractères, au lieu d’appeler `re.compile` avec le flag `re.I`, vous pouvez utiliser `(?i)` comme ceci :

```
import re
```

```
# on peut embarquer les flags comme IGNORECASE
# directement dans la regexp
# c'est équivalent de faire ceci

re_obj = re.compile("abc", flags=re.IGNORECASE)
re_obj.match("ABC").group(0)
```

```
'ABC'
```

```
# ou cela

re.match("(?i)abc", "ABC").group(0)
```

```
'ABC'
```

```
# les flags comme (?i) doivent apparaître
# en premier dans la regexp
re.match("abc(?i)", "ABC").group(0)
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:3:
↳ DeprecationWarning: Flags not at the start of the expression 'abc(?i)'
This is separate from the ipykernel package so we can avoid doing imports until
```

```
'ABC'
```

Pour plus de précisions sur ce trait, que nous avons laissé de coté dans le complément pour ne pas trop l’alourdir, voyez [la documentation sur les expressions régulières](#) et cherchez la première occurrence de `ilmsux`.

Décortiquer une URL

On vous demande d’écrire une expression régulière qui permette d’analyser des URLs.

Voici les conventions que nous avons adoptées pour l’exercice :

- la chaîne contient les parties suivantes

- `<protocol>://<location>/<path>`
- l'url commence par le nom d'un protocole qui doit être parmi `http, https, ftp, ssh`
- le nom du protocole peut contenir de manière indifférente des minuscules ou des majuscules,
- ensuite doit venir la séquence `://`
- ensuite on va trouver une chaîne `<location>` qui contient :
 - potentiellement un nom d'utilisateur, et s'il est présent, potentiellement un mot de passe, * obligatoirement un nom de `hostname`, * potentiellement un numéro de port;
- lorsque les 4 parties sont présentes dans `<location>`, cela se présente comme ceci :
 - `<location> = <user>:<password>@<hostname>:<port>`
- si l'on note entre crochets les parties optionnelles, cela donne :
 - `<location> = [<user>[:<password>]@]<hostname>[:<port>]`
- le champ `<user>` ne peut contenir que des caractères alphanumériques; si le `@` est présent le champ `<user>` ne peut pas être vide
- le champ `<password>` peut contenir tout sauf un `:` et de même, si le `:` est présent le champ `<password>` ne peut pas être vide * le champ `<hostname>` peut contenir une suite non-vide de caractères alphanumériques, underscores, ou `.` * le champ `<port>` ne contient que des chiffres, et il est non vide si le `:` est spécifié * le champ `<path>` peut être vide.

Enfin, vous devez définir les groupes `proto`, `user`, `password`, `hostname`, `port` et `path` qui sont utilisés pour vérifier votre résultat. Dans la case **Résultat attendu**, vous trouverez soit `None` si la regexp ne filtre pas l'intégralité de l'entrée, ou bien une liste ordonnée de tuples qui donnent la valeur de ces groupes; vous n'avez rien à faire pour construire ces tuples, c'est l'exercice qui s'en occupe.

```
# pour charger l'exercice
from corrections.regex_url import exo_url
```

```
# exemples du résultat attendu
exo_url.example()
```

```
# n'hésitez pas à construire votre regexp petit à petit

regex_url = "<votre_regex>"
```

```
exo_url.correction(regex_url)
```

Les slices en python

28.1 Complément - niveau basique

Ce support de cours reprend les notions de *slicing* vues dans la vidéo.

Nous allons illustrer les slices sur la chaine suivante, rappelez-vous toutefois que ce mécanisme fonctionne avec toutes les séquences que l'on verra plus tard, comme les listes ou les tuples.

```
chaine = "abcdefghijklmnopqrstuvwxyz" ; print(chaine)
```

```
abcdefghijklmnopqrstuvwxyz
```

28.1.1 Slice sans pas

On a vu en cours qu'une slice permet de désigner toute une plage d'éléments d'une séquence. Ainsi on peut écrire

```
chaine[2:6]
```

```
'cdef'
```

28.1.2 Conventions de début et fin

Les débutants ont parfois du mal avec les bornes. Il faut se souvenir que

- les indices **commencent** comme toujours à **zéro**
- le premier indice **debut** est **inclus**
- le second indice **fin** est **exclu**
- on obtient en tout **fin-debut** items dans le résultat

Ainsi ci-dessus le résultat contient $6 - 2 = 4$ éléments.

Pour vous aider à vous souvenir des conventions de début et de fin, souvenez-vous qu'on veut pouvoir facilement juxtaposer deux slices qui ont une borne commune.

C'est-à-dire qu'avec

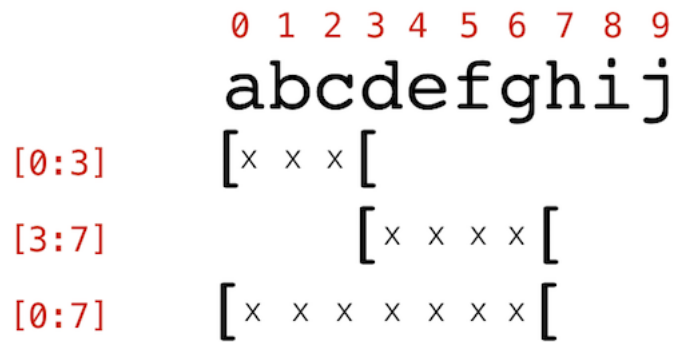


Fig. 28.1 – début et fin

```
# chaine[a:b] + chaine[b:c] == chaine[a:c]
chaine[0:3] + chaine[3:7] == chaine[0:7]
```

True

Bornes omises

On peut omettre une borne

```
# si on omet la première borne, cela signifie que
# la slice commence au début de l'objet
chaine[:6]
```

'abcdef'

```
# et bien entendu c'est la même chose si on omet la deuxième borne
chaine[24:]
```

'yz'

```
# ou même omettre les deux bornes, auquel cas on
# fait une copie de l'objet - on y reviendra plus tard
chaine[:]
```

'abcdefghijklmnopqrstuvwxyz'

Indices négatifs

On peut utiliser des indices négatifs pour compter à partir de la fin

```
chaine[3:-3]
```

'defghijklmnopqrstuvw'

```
chaine[-3:]
```

'xyz'

28.1.3 Slice avec pas

Il est également possible de préciser un *pas*, de façon à ne choisir par exemple, dans la plage donnée, qu'un élément sur deux

```
# le pas est précisé après un deuxième deux-points (:)
# ici on va choisir un caractère sur deux dans la plage [3:-3]
chaine[3:-3:2]
```

```
'dfhjlnprtv'
```

Comme on le devine, le troisième élément de la slice, ici 2, détermine le pas. On ne retient donc, dans la chaîne `defghi...` que `d`, puis `f`, et ainsi de suite.

On peut préciser du coup la borne de fin (ici -3) avec un peu de liberté, puisqu'ici on obtiendrait un résultat identique avec -4.

```
chaine[3:-4:2]
```

```
'dfhjlnprtv'
```

28.1.4 Pas négatif

Il est même possible de spécifier un pas négatif. Dans ce cas, de manière un peu contre-intuitive, il faut préciser un début (le premier indice de la slice) qui soit *plus à droite* que la fin (le second indice).

Pour prendre un exemple, comme l'élément d'indice -3, c'est-à-dire `x`, est plus à droite que l'élément d'indice 3, c'est-à-dire `d`, évidemment si on ne précisait pas le pas (qui revient à choisir un pas égal à 1), on obtiendrait une liste vide.

```
chaine[-3:3]
```

```
''
```

Si maintenant on précise un pas négatif, on obtient cette fois

```
chaine[-3:3:-2]
```

```
'xvtrpnljhf'
```

28.1.5 Conclusion

À nouveau, souvenez-vous que tous ces mécanismes fonctionnent avec de nombreux autres types que les chaînes de caractères. En voici deux exemples qui anticipent tous les deux sur la suite, mais qui devraient illustrer les vastes possibilités qui sont offertes avec les slices.

Listes

Par exemple sur les listes

```
liste = [0, 2, 4, 8, 16, 32, 64, 128]
liste
```

```
[0, 2, 4, 8, 16, 32, 64, 128]
```

```
liste[-1:1:-2]
```

```
[128, 32, 8]
```

Et même ceci, qui peut être déroutant. Nous reviendrons dessus.

```
liste[2:4] = [100, 200, 300, 400, 500]
liste
```

```
[0, 2, 100, 200, 300, 400, 500, 16, 32, 64, 128]
```

28.2 Complément - niveau avancé

La librairie `numpy` permet de manipuler des tableaux ou matrices. En anticipant (beaucoup) sur son usage que nous reverrons bien entendu en détails, voici un aperçu de ce qu'on peut faire avec des slices sur des objets `numpy`.

```
# ces deux premières cellules sont à admettre
# on construit un tableau ligne
import numpy as np

un_cinq = np.array([1, 2, 3, 4, 5])
un_cinq
```

```
array([1, 2, 3, 4, 5])
```

```
# ces deux premières cellules sont à admettre
# on le combine avec lui-même - et en utilisant une slice un peu magique
# pour former un tableau carré 5x5

array = 10 * un_cinq[:, np.newaxis] + un_cinq
array
```

```
array([[11, 12, 13, 14, 15],
       [21, 22, 23, 24, 25],
       [31, 32, 33, 34, 35],
       [41, 42, 43, 44, 45],
       [51, 52, 53, 54, 55]])
```

Sur ce tableau de taille 5x5, nous pouvons aussi faire du slicing et extraire le sous-tableau 3x3 au centre

```
centre = array[1:4, 1:4]
centre
```

```
array([[22, 23, 24],
       [32, 33, 34],
       [42, 43, 44]])
```

On peut bien sûr également utiliser un pas

```
coins = array[::4, ::4]
coins
```

```
array([[11, 15],
       [51, 55]])
```

Ou bien retourner complètement dans une direction

```
tete_en_bas = array[::-1,:]  
tete_en_bas
```

```
array([[51, 52, 53, 54, 55],  
       [41, 42, 43, 44, 45],  
       [31, 32, 33, 34, 35],  
       [21, 22, 23, 24, 25],  
       [11, 12, 13, 14, 15]])
```

Méthodes spécifiques aux listes

29.1 Complément - niveau basique

Voici quelques unes des méthodes disponibles sur le type `list`.

29.1.1 Trouver l'information

Pour commencer, rappelons comment retrouver la liste des méthodes définies sur le type `list`.

```
help(list)
```

Ignorez les méthodes dont le nom commence et termine par `__` (nous parlerons de ceci en semaine 6), vous trouvez alors les méthodes utiles listées entre `append` et `sort`.

Certaines de ces méthodes ont été vues dans la vidéo sur les séquences, c'est le cas notamment de `count` et `index`.

Nous allons à présent décrire les autres, partiellement et brièvement. Un autre complément décrit la méthode `sort`. Reportez-vous au lien donné en fin de notebook pour obtenir une information plus complète.

Donnons-nous pour commencer une liste témoin.

```
liste = [0, 1, 2, 3]
print('liste', liste)
```

```
liste [0, 1, 2, 3]
```

Avertissements :

- soyez bien attentifs au nombre de fois où vous exécutez les cellules de ce notebook
- par exemple une liste renversée deux fois peut donner l'impression que `reverse` ne marche pas :)
- n'hésitez pas à utiliser le menu *Cell -> Run All* pour réexécuter en une seule fois le notebook entier.

29.1.2 append

La méthode `append` permet d'ajouter **un élément** à la fin d'une liste :

```
liste.append('ap')
print('liste', liste)
```

```
liste [0, 1, 2, 3, 'ap']
```

29.1.3 extend

La méthode `extend` réalise la même opération, mais avec **tous les éléments** de la liste qu'on lui passe en argument :

```
liste2 = ['ex1', 'ex2']
liste.extend(liste2)
print('liste', liste)
```

```
liste [0, 1, 2, 3, 'ap', 'ex1', 'ex2']
```

29.1.4 append vs +

Ces deux méthodes `append` et `extend` sont donc assez voisines; avant de voir d'autres méthodes de `list`, prenons un peu le temps de comparer leur comportement avec l'addition `+` de liste. L'élément clé ici, on l'a déjà vu dans la vidéo, est que la liste est un objet **mutable**. `append` et `extend` **modifient** la liste sur laquelle elles travaillent, alors que l'addition **crée un nouvel objet**.

```
# pour créer une liste avec les n premiers entiers, on utilise
# la fonction built-in range(), que l'on convertit en liste
# on aura l'occasion d'y revenir
a1 = list(range(3))
print(a1)
```

```
[0, 1, 2]
```

```
a2 = list(range(10, 13))
print(a2)
```

```
[10, 11, 12]
```

```
# le fait d'utiliser + crée une nouvelle liste
a3 = a1 + a2
```

```
# si bien que maintenant on a trois objets différents
print('a1', a1)
print('a2', a2)
print('a3', a3)
```

```
a1 [0, 1, 2]
a2 [10, 11, 12]
a3 [0, 1, 2, 10, 11, 12]
```

Comme on le voit, après une addition, les deux termes de l'addition sont inchangés. Pour bien comprendre, voyons exactement le même scénario sous `pythontutor` :


```
%load_ext ipythontutor
```

Note : une fois que vous avez évalué la cellule avec `%%ipythontutor`, vous devez cliquer sur le bouton Forward pour voir pas à pas le comportement du programme.

```
%%ipythontutor height=230 ratio=0.7
a1 = list(range(3))
a2 = list(range(10, 13))
a3 = a1 + a2
```

Alors que si on avait utilisé `extend`, on aurait obtenu ceci :

```
%%ipythontutor height=200 ratio=0.75
e1 = list(range(3))
e2 = list(range(10, 13))
e3 = e1.extend(e2)
```

Ici on tire profit du fait que la liste est un objet mutable : `extend` **modifie** l'objet sur lequel on l'appelle (ici `e1`). Dans ce scénario on ne crée en tout que deux objets, et du coup il est inutile pour `extend` de renvoyer quoi que ce soit, et c'est pourquoi `e3` ici vaut `None`.

C'est pour cette raison que :

- l'addition est disponible sur tous les types séquences - on peut toujours réaliser l'addition puisqu'on crée un nouvel objet pour stocker le résultat de l'addition ;
- mais `append` et `extend` ne sont par exemple **pas disponibles** sur les chaînes de caractères, qui sont **immuables** - si `e1` était une chaîne, on ne pourrait pas la modifier pour lui ajouter des éléments.

29.1.5 insert

Reprenons notre inventaire des méthodes de `list`, et pour cela rappelons nous le contenu de la variable `liste` :

```
liste
```

```
[0, 1, 2, 3, 'ap', 'ex1', 'ex2']
```

La méthode `insert` permet, comme le nom le suggère, d'insérer un élément à une certaine position ; comme toujours les indices commencent à zéro et donc :

```
# insérer à l'index 2
liste.insert(2, '1 bis')
print('liste', liste)
```

```
liste [0, 1, '1 bis', 2, 3, 'ap', 'ex1', 'ex2']
```

On peut remarquer qu'un résultat analogue peut être obtenu avec une affectation de slice ; par exemple pour insérer au rang 5 (i.e. avant `ap`), on pourrait aussi bien faire :

```
liste[5:5] = ['3 bis']
print('liste', liste)
```

```
liste [0, 1, '1 bis', 2, 3, '3 bis', 'ap', 'ex1', 'ex2']
```

29.1.6 remove

La méthode `remove` détruit la **première occurrence** d'un objet dans la liste :

```
liste.remove(3)
print('liste', liste)
```

```
liste [0, 1, '1 bis', 2, '3 bis', 'ap', 'ex1', 'ex2']
```

29.1.7 pop

La méthode `pop` prend en argument un indice ; elle permet d'extraire l'élément à cet indice. En un seul appel on obtient la valeur de l'élément et on l'enlève de la liste :

```
popped = liste.pop(0)
print('popped', popped, 'liste', liste)
```

```
popped 0 liste [1, '1 bis', 2, '3 bis', 'ap', 'ex1', 'ex2']
```

Si l'indice n'est pas précisé, c'est le dernier élément de la liste qui est visé

```
popped = liste.pop()
print('popped', popped, 'liste', liste)
```

```
popped ex2 liste [1, '1 bis', 2, '3 bis', 'ap', 'ex1']
```

29.1.8 reverse

Enfin `reverse` renverse la liste, le premier élément devient le dernier :

```
liste.reverse()
print('liste', liste)
```

```
liste ['ex1', 'ap', '3 bis', 2, '1 bis', 1]
```

On peut remarquer ici que le résultat se rapproche de ce qu'on peut obtenir avec une opération de slicing comme ceci

```
liste2 = liste[::-1]
print('liste2', liste2)
```

```
liste2 [1, '1 bis', 2, '3 bis', 'ap', 'ex1']
```

à la différence toutefois qu'avec le slicing c'est une copie de la liste initiale qui est retournée, la liste de départ n'est quant à elle pas modifiée.

29.1.9 Pour en savoir plus

<https://docs.python.org/3/tutorial/datastructures.html#more-on-lists>

29.1.10 Note spécifique aux notebooks

help avec ?

Je vous signale en passant que dans un notebook vous pouvez obtenir de l'aide avec un point d'interrogation ? inséré avant ou après un symbole. Par exemple pour obtenir des précisions sur la méthode `list.pop`, on peut faire soit :

```
# fonctionne dans tous les environnements python
help(list.pop)
```

Help on method_descriptor:

```
pop(...)
    L.pop([index]) -> item -- remove and return item at index (default last).
    Raises IndexError if list is empty or index is out of range.
```

```
# spécifique aux notebooks
# l'affichage obtenu est légèrement différent
# tapez la touche 'Esc' - ou cliquez la petite croix
# pour faire disparaître le dialogue qui apparaît en bas
list.pop?
```

Complétion avec Tab

Dans un notebook vous avez aussi la complétion ; si vous tapez - dans une cellule de code - le début d'un symbole connu dans l'environnement :

```
# placez votre curseur à la fin de la ligne après 'li'
# et appuyez sur la touche 'Tab'
license
```

```
Type license() to see the full license text
```

Vous voyez apparaître un dialogue avec les noms connus qui commencent par `li` ; utilisez les flèches pour choisir, et “Return” pour sélectionner.

Objets mutables et objets immuables

30.1 Complément - niveau basique

30.1.1 Les chaînes sont des objets immuables

Voici un exemple d'un fragment de code qui illustre le caractère immuable des chaînes de caractères. Nous l'exécutons sous `pythontutor`, afin de bien illustrer les relations entre variables et objets.

```
# il vous faut charger cette cellule
# pour pouvoir utiliser les suivantes
%load_ext ipythontutor
```

Note : une fois que vous avez évalué la cellule avec `%%ipythontutor`, vous devez cliquer sur le bouton Forward pour voir pas à pas le comportement du programme.

Le scénario est très simple, on crée deux variables `s1` et `s2` vers le même objet `'abc'`, puis on fait une opération `+=` sur la variable `s1`.

Comme l'objet est une chaîne, il est donc immuable, on ne **peut pas modifier l'objet** directement ; pour obtenir l'effet recherché (à savoir que `s1` s'allonge de `'def'`), python **crée un deuxième objet**, comme on le voit bien sous `pythontutor` :

```
%%ipythontutor heapPrimitives=true
# deux variables vers le même objet
s1 = 'abc'
s2 = s1
# on essaie de modifier l'objet
s1 += 'def'
# pensez à cliquer sur `Forward`
```

30.1.2 Les listes sont des objets mutables

Voici ce qu'on obtient par contraste pour le même scénario mais qui cette fois utilise des listes, qui sont des objets mutables :

```
%%ipythontutor heapPrimitives=true ratio=0.8
# deux variables vers le même objet
```

```
listel = ['a', 'b', 'c']
liste2 = listel
# on modifie l'objet
listel += ['d', 'e', 'f']
# pensez à cliquer sur `Forward`
```

30.1.3 Conclusion

Ce comportement n'est pas propre à l'usage de l'opérateur += - que pour cette raison d'ailleurs nous avons tendance à déconseiller.

Les objets mutables et immuables ont par essence un comportement différent, il est très important d'avoir ceci présent à l'esprit.

Nous aurons notamment l'occasion d'approfondir cela dans la séquence consacrée aux références partagées, en semaine 3.

31.1 Complément - niveau basique

Python fournit une méthode standard pour trier une liste, qui s'appelle, sans grande surprise, `sort`.

31.1.1 La méthode `sort`

Voyons comment se comporte `sort` sur un exemple simple.

```
liste = [8, 7, 4, 3, 2, 9, 1, 5, 6]
print('avant tri', liste)
liste.sort()
print('apres tri', liste)
```

```
avant tri [8, 7, 4, 3, 2, 9, 1, 5, 6]
apres tri [1, 2, 3, 4, 5, 6, 7, 8, 9]
```

On retrouve ici, avec l'instruction `liste.sort()` un cas d'appel de méthode (ici `sort`) sur un objet (ici `liste`), comme on l'avait vu dans la vidéo sur la notion d'objet.

La première chose à remarquer est que la liste d'entrée a été modifiée, on dit « en place », ou encore « par effet de bord ». Voyons cela sous `pythontutor` :

```
%load_ext ipythontutor
```

```
%%ipythontutor height=200 ratio=0.8
liste = [3, 2, 9, 1]
liste.sort()
```

On aurait pu imaginer que la liste d'entrée soit restée inchangée, et que la méthode de tri renvoie une copie triée de la liste, ce n'est pas le choix qui a été fait, cela permet d'économiser des allocations mémoire autant que possible et d'accélérer sensiblement le tri.

31.1.2 La fonction `sorted`

Si vous avez besoin de faire le tri sur une copie de votre liste, la fonction `sorted` vous permet de le faire :

```
%ipythontutor height=200 ratio=0.8
liste1 = [3, 2, 9, 1]
liste2 = sorted(liste1)
```

31.1.3 Tri décroissant

Revenons à la méthode `sort` et aux tris *en place*. Par défaut la liste est triée par ordre croissant, si au contraire vous voulez l'ordre décroissant, faites comme ceci :

```
liste = [8, 7, 4, 3, 2, 9, 1, 5, 6]
print('avant tri', liste)
liste.sort(reverse=True)
print('après tri décroissant', liste)
```

```
avant tri [8, 7, 4, 3, 2, 9, 1, 5, 6]
après tri décroissant [9, 8, 7, 6, 5, 4, 3, 2, 1]
```

Nous n'avons pas encore vu à quoi correspond cette formule `reverse=True` dans l'appel à la méthode - ceci sera approfondi dans le chapitre sur les appels de fonction - mais dans l'immédiat vous pouvez utiliser cette technique telle quelle.

31.1.4 Chaînes de caractères

Cette technique fonctionne très bien sur tous les types numériques (enfin, à l'exception des complexes ; en guise d'exercice : pourquoi ?), ainsi que sur les chaînes de caractères :

```
liste = ['spam', 'egg', 'bacon', 'beef']
liste.sort()
print('après tri', liste)
```

```
après tri ['bacon', 'beef', 'egg', 'spam']
```

Comme on s'y attend, il s'agit cette fois d'un **tri lexicographique**, dérivé de l'ordre sur les caractères. Autrement dit, c'est l'ordre du dictionnaire. Il faut souligner toutefois, pour les personnes n'ayant jamais été exposées à l'informatique, que cet ordre, quoique déterministe, est arbitraire en dehors des lettres de l'alphabet.

Ainsi par exemple :

```
# deux caractères minuscules se comparent
# comme on s'y attend
'a' < 'z'
```

True

Bon, mais par contre :

```
# si l'un est en minuscule et l'autre en majuscule,
# ce n'est plus le cas
'Z' < 'a'
```

True

Ce qui à son tour explique ceci :

```
# la conséquence de 'Z' < 'a', c'est que
liste = ['abc', 'Zoo']
liste.sort()
print(liste)
```



```
['Zoo', 'abc']
```

Et lorsque les chaînes contiennent des espaces ou autres ponctuations, le résultat du tri peut paraître surprenant :

```
# attention ici notre première chaîne commence par un espace
# et le caractère 'Espace' est plus petit
# que tous les autres caractères imprimables
liste = [' zoo', 'ane']
liste.sort()
print(liste)
```

```
[' zoo', 'ane']
```

31.1.5 À suivre

Il est possible de définir soi-même le critère à utiliser pour trier une liste, et nous verrons cela bientôt, une fois que nous aurons introduit la notion de fonction.

Indentations en python

32.1 Complément - niveau basique

32.1.1 Imbrications

Nous l'avons vu dans la vidéo, la pratique la plus courante est d'utiliser systématiquement une indentation de 4 espaces :

```
# la convention la plus généralement utilisée
# consiste à utiliser une indentation de 4 espaces
if 'g' in 'egg':
    print('OUI')
else:
    print('NON')
```

OUI

Voyons tout de suite comment on pourrait écrire plusieurs tests imbriqués :

```
entree = 'spam'

# pour imbriquer il suffit d'indenter de 8 espaces
if 'a' in entree:
    if 'b' in entree:
        cas11 = True
        print('a et b')
    else:
        cas12 = True
        print('a mais pas b')
else:
    if 'b' in entree:
        cas21 = True
        print('b mais pas a')
    else:
        cas22 = True
        print('ni a ni b')
```

```
a mais pas b
```

Dans cette construction assez simple, remarquez bien **les deux points “:”** à chaque début de bloc, c’est-à-dire à chaque fin de ligne `if` ou `else`.

Cette façon d’organiser le code peut paraître très étrange, notamment aux gens habitués à un autre langage de programmation, puisqu’en général les syntaxes des langages sont conçues de manière à être insensibles aux espaces et à la présentation.

Comme vous le constaterez à l’usage cependant, une fois qu’on s’y est habitué cette pratique est très agréable, une fois qu’on a écrit la dernière ligne du code, on n’a pas à réfléchir à re fermer le bon nombre d’accolades ou de *end*.

Par ailleurs, comme pour tous les langages, votre éditeur favori connaît cette syntaxe et va vous aider à respecter la règle des 4 caractères. Nous ne pouvons pas publier ici une liste des commandes disponibles par éditeur, nous vous invitons le cas échéant à échanger entre vous sur le forum pour partager les recettes que vous utilisez avec votre éditeur / environnement de programmation favori.

32.2 Complément - niveau intermédiaire

32.2.1 Espaces vs tabulations

Version courte

Il nous faut par contre donner quelques détails sur un problème que l’on rencontre fréquemment sur du code partagé entre plusieurs personnes quand celles-ci utilisent des environnements différents.

Pour faire court, ce problème est **susceptible d’apparaître dès qu’on utilise des tabulations**, plutôt que des espaces, pour implémenter les indentations. Aussi, le message à retenir ici est **de ne jamais utiliser de tabulations dans votre code python**. Tout bon éditeur python devrait faire cela par défaut.

Version longue

En version longue, il existe un code ASCII pour un caractère qui s’appelle *Tabulation* (alias Control-i, qu’on note aussi $\wedge I$); l’interprétation de ce caractère n’étant pas clairement spécifiée, il arrive qu’on se retrouve dans une situation comme la suivante.

Bernard utilise l’éditeur `vim`; sous cet éditeur il lui est possible de mettre des tabulations dans son code, et de choisir la valeur de ces tabulations. Aussi il va dans les préférences de `vim`, choisit `Tabulation=4`, et écrit un programme qu’il voit comme ceci

```
if 'a' in entree:
    if 'b' in entree:
        cas11 = True
        print('a et b')
    else:
        cas12 = True
        print('a mais pas b')
```

```
a mais pas b
```

Sauf qu’en fait, il a mis un mélange de tabulations et d’espaces, et en fait le fichier contient (avec $\wedge I$ pour tabulation) :

```
if 'a' in entree:
^Iif 'b' in entree:
^I^Icas11 = True
    ^Iprint 'a et b'
^Ielse:
```

```
^I^Icas12 = True
    ^Iprint 'a mais pas b'
```

Remarquez le mélange de Tabulations et d'espaces dans les deux lignes avec `print`. Bernard envoie son code à Alice qui utilise `emacs`. Dans son environnement, `emacs` affiche une tabulation comme 8 caractères. Du coup Alice « voit » le code suivant

```
if 'a' in entree:
    if 'b' in entree:
        cas11 = True
        print 'a et b'
    else:
        cas12 = True
        print 'a mais pas b'
```

Bref, c'est la confusion la plus totale. Aussi répétons-le, **n'utilisez jamais de tabulations dans votre code python**.

Ce qui ne veut pas dire qu'il ne faut pas utiliser la touche `Tab` avec votre éditeur - au contraire, c'est une touche très utilisée - mais faites bien la différence entre le fait d'appuyer sur la touche `Tab` et le fait que le fichier sauvé sur disque contient effectivement un caractère tabulation. Votre éditeur favori propose très certainement une option permettant de faire les remplacements idoines pour ne pas écrire de tabulation dans vos fichiers, tout en vous permettant d'indenter votre code avec la touche `Tab`.

Signalons enfin que `python3` est plus restrictif que `python2` à cet égard, et interdit de mélanger des espaces et des tabulations sur une même ligne. Ce qui n'enlève rien à notre recommandation.

32.3 Complément - niveau avancé

Vous pouvez trouver du code qui ne respecte pas la convention des 4 caractères.

En version courte : **Utilisez toujours des indentations de 4 espaces**.

En version longue, et pour les curieux : `python` **n'impose pas** que les indentations soient de 4 caractères. Aussi vous pouvez rencontrer un code qui ne respecte pas cette convention, et il nous faut, pour être tout à fait précis sur ce que `python` accepte ou non, préciser ce qui est réellement requis par `python`.

La règle utilisée pour analyser votre code, c'est que toutes les instructions **dans un même bloc** sont présentées avec le même niveau d'indentation. Si deux lignes successives - modulo les blocs imbriqués - ont la même indentation, elles sont dans le même bloc.

Voyons quelques exemples. Tout d'abord le code suivant est **légal**, quoique, redisons-le pour la dernière fois, **pas du tout recommandé** :

```
# code accepté mais pas du tout recommandé
if 'a' in 'pas du tout recommande':
    succes = True
    print('OUI')
else:
    print('NON')
```

```
OUI
```

En effet les deux blocs (après `if` et après `else`) sont des blocs distincts, ils sont libres d'utiliser deux indentations différentes (ici 2 et 6)

Par contre la construction ci-dessous n'est pas légale

```
# ceci n'est pas correct et rejeté par python
if 'a' in entree:
    if 'b' in entree:
```

```
cas11 = True
print 'a et b'
else:
    cas12 = True
    print 'a mais pas b'
```

```
File "<tokenize>", line 6
    else:
    ^
IndentationError: unindent does not match any outer indentation level
```

En effet les deux lignes `if` et `else` font logiquement partie du même bloc, elles **doivent** donc avoir la même indentation. Avec cette présentation le lecteur python émet une erreur et ne peut pas interpréter le code.

Bonnes pratiques de présentation de code

33.1 Complément - niveau basique

33.1.1 La PEP-008

On trouve [dans la PEP-008 \(en anglais\)](#) les conventions de codage qui s'appliquent à toute la librairie standard, et qui sont certainement un bon point de départ pour vous aider à trouver le style de présentation qui vous convient.

Nous vous recommandons en particulier les sections sur

- l'indentation
- les espaces
- les commentaires

33.1.2 Un peu de lecture : le module `pprint`

Voici par exemple le code du module `pprint` (comme `PrettyPrint`) de la librairie standard qui permet d'imprimer des données.

La fonction du module - le pretty printing - est évidemment accessoire ici, mais vous pouvez y voir illustré

- le *docstring* pour le module : les lignes de 11 à 35,
- les indentations, comme nous l'avons déjà mentionné sont à 4 espaces, et sans tabulation,
- l'utilisation des espaces, notamment autour des affectations et opérateurs, des définitions de fonction, des appels de fonctions...
- les lignes qui restent dans une largeur « raisonnable » (79 caractères) * vous pouvez regarder notamment la façon de couper les lignes pour respecter cette limite en largeur.

```
from modtools import show_module_html
import pprint
show_module_html(pprint, lineno_width=3)
```

33.1.3 Espaces

Comme vous pouvez le voir dans `pprint.py`, les règles principales concernant les espaces sont les suivantes.

- S'agissant des **affectations** et **opérateurs**, on fera

```
x = y + z
```

Et non pas

```
~~`x=y+z`~~
```

Ni

```
~~`x = y+z`~~
```

Ni encore

```
~~`x=y + z`~~
```

L'idée étant d'aérer de manière homogène pour faciliter la lecture.

- On **déclare une fonction** comme ceci

```
def foo(x, y, z):
```

Et non pas comme ceci (un espace en trop avant la parenthèse ouvrante)

```
~~`def foo (x, y, z):`~~
```

Ni surtout comme ceci (pas d'espace entre les paramètres)

```
~~`def foo (x,y,z):`~~
```

- La même règle s'applique naturellement aux **appels de fonction** :

```
foo(x, y, z)
```

et non pas ~~[STRIKEOUT]~~ `foo (x, y, z)`

ni ~~[STRIKEOUT]~~ `def foo (x, y, z):`

Il est important de noter qu'il s'agit ici de **règles d'usage** et non pas de règles syntaxiques ; tous les exemples barrés ci-dessus sont en fait **syntactiquement corrects**, l'interpréteur les accepterait sans souci ; mais ces règles sont **très largement adoptées**, et obligatoires pour intégrer du code dans la librairie standard.

33.1.4 Coupures de ligne

Nous allons à présent zoomer dans ce module pour voir quelques exemples de coupure de ligne. Par contraste avec ce qui précède, il s'agit cette fois surtout de **règles syntaxiques**, qui peuvent rendre un code non valide si elles ne sont pas suivies.

Coupure de ligne sans *backslash* (`\`)

```
show_module_html(pprint,
                  beg="def pprint",
                  end="def pformat",
                  lineno_width=3)
```

La fonction `pprint` (ligne ~47) est une commodité (qui crée une instance de `PrettyPrinter`, sur lequel on envoie la méthode `pprint`).

Vous voyez ici qu'il n'est **pas nécessaire** d'insérer un *backslash* (\) à la fin des lignes 50 et 51, car il y a une parenthèse ouvrante qui n'est pas fermée à ce stade.

De manière générale, lorsqu'une parenthèse ouvrante (- idem avec les crochets [et accolades { - n'est pas fermée sur la même ligne, l'interpréteur suppose qu'elle sera fermée plus loin et n'impose pas de *backslash*.

Ainsi par exemple on peut écrire sans *backslash* :

```
valeurs = [
    1,
    2,
    3,
    5,
    7,
]
```

Ou encore

```
x = un_nom_de_fonction_tres_tres_long(
    argument1, argument2,
    argument3, argument4,
)
```

À titre de rappel, signalons aussi les chaînes de caractères à base de `"""` ou `'''` qui permettent elles aussi d'utiliser plusieurs lignes consécutives sans *backslash*, comme

```
texte = """ Les sanglots longs
Des violons
De l'automne"""
```

Coupure de ligne avec *backslash* (\)

Par contre il est des cas où le *backslash* est nécessaire :

```
show_module_html(pprint,
                  beg="components), readable, recursive",
                  end="elif len(object) ",
                  lineno_width=3)
```

Dans ce fragment au contraire, vous voyez en ligne 522 qu'**il a fallu cette fois** insérer un *backslash* \ comme caractère de continuation pour que l'instruction puisse se poursuivre en ligne 523.

Coupages de lignes - épilogue

Dans tous le cas où une instruction est répartie sur plusieurs lignes, c'est naturellement l'indentation de **la première ligne** qui est significative pour savoir à quel bloc rattacher cette instruction.

Notez bien enfin qu'on peut toujours mettre un *backslash* même lorsque ce n'est pas nécessaire, mais on évite cette pratique en règle générale car les *backslash* nuisent à la lisibilité.

33.2 Complément - niveau intermédiaire

33.2.1 Outils liés à PEP008

Il existe plusieurs outils liés à la PEP0008, pour vérifier si votre code est conforme, ou même le modifier pour qu'il le devienne.

Ce qui nous donne un excellent prétexte pour parler un peu de <https://pypi.python.org>, qui est la plateforme qui distribue les logiciels disponibles via l'outil `pip3`.

Je vous signale notamment :

- <https://pypi.python.org/pypi/pep8/> pour vérifier, et
- <https://pypi.python.org/pypi/autopep8/> pour modifier automatiquement votre code et le rendre conforme.

33.2.2 Les deux-points “:”

Dans un autre registre entièrement, vous pouvez [vous reporter à ce lien](#) si vous êtes intéressé par la question de savoir pourquoi on a choisi un délimiteur (le caractère deux-points :) pour terminer les instructions comme `if`, `for` et `def`.

34.1 Complément - niveau basique

Nous avons vu qu'en python les blocs de code sont définis par leur indentation.

34.1.1 Une fonction vide

Cette convention a une limitation lorsqu'on essaie de définir un bloc vide. Voyons par exemple comment on définirait en C une fonction qui ne fait rien

```
/* une fonction C qui ne fait rien */  
void foo () {}
```

Comme en python on n'a pas d'accolade pour délimiter les blocs de code, il existe une instruction `pass`, qui ne fait rien. À l'aide de cette instruction on peut à présent définir une fonction vide comme ceci :

```
# une fonction python qui ne fait rien  
def foo():  
    pass
```

34.1.2 Une boucle vide

Pour prendre un second exemple un peu plus pratique, et pour anticiper un peu sur l'instruction `while` que nous verrons très bientôt, voici un exemple d'une boucle vide, c'est à dire sans corps, qui permet de « dépiler » dans une liste jusqu'à l'obtention d'une certaine valeur :

```
liste = list(range(10))  
print('avant', liste)  
while liste.pop() != 5:  
    pass  
print('après', liste)
```

```
avant [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]  
après [0, 1, 2, 3, 4]
```

On voit qu'ici encore l'instruction `pass` a toute son utilité.

34.2 Complément - niveau intermédiaire

34.2.1 Un if sans then

```
# on utilise dans ces exemples une condition fausse
condition = False
```

Imaginons qu'on parte d'un code hypothétique qui fasse ceci :

```
# la version initiale
if condition:
    print("non")
else:
    print("bingo")
```

```
bingo
```

et que l'on veuille modifier ce code pour simplement supprimer l'impression de non. La syntaxe du langage **ne permet pas** de simplement commenter le premier print :

```
# si on commente le premier print
# la syntaxe devient incorrecte
if condition:
#     print "non"
else:
    print "bingo"
```

```
File "<ipython-input-5-4167a77e3cac>", line 5
    else:
    ^
IndentationError: expected an indented block
```

Évidemment ceci pourrait être réécrit autrement en inversant la condition, mais parfois on s'efforce de limiter au maximum l'impact d'une modification sur le code. Dans ce genre de situation on préférera écrire plutôt

```
# on peut s'en sortir en ajoutant une instruction pass
if condition:
#     print "non"
    pass
else:
    print("bingo")
```

```
bingo
```

34.2.2 Une classe vide

Enfin comme on vient de le voir dans la vidéo, on peut aussi utiliser `pass` pour définir une classe vide comme ceci :

```
class Foo:
    pass
```

```
foo = Foo()
```

Fonctions avec ou sans valeur de retour

35.1 Complément - niveau basique

35.1.1 Le style procédural

Une procédure est une fonction qui se contente de dérouler des instructions. Voici un exemple d'une telle fonction :

```
def affiche_carre(n):  
    print("le carre de", n, "vaut", n*n)
```

qui s'utiliserait comme ceci

```
affiche_carre(12)
```

```
le carre de 12 vaut 144
```

35.1.2 Le style fonctionnel

Mais en fait, il serait dans notre cas beaucoup plus commode de définir une fonction qui **retourne** le carré d'un nombre, afin de pouvoir écrire quelque chose comme :

```
surface = carre(15)
```

quitte à imprimer cette valeur ensuite si nécessaire. Jusqu'ici nous avons fait beaucoup appel à `print`, mais dans la pratique, imprimer n'est pas un but en soi, au contraire bien souvent.

35.1.3 L'instruction `return`

Voici comment on pourrait écrire une fonction `carre` qui **retourne** (on dit aussi **renvoie**) le carré de son argument :

```
def carre(n):  
    return n*n
```

```
if carre(8) <= 100:
    print('petit appartement')
```

```
petit appartement
```

La sémantique (le mot savant pour « comportement ») de l’instruction `return` est assez simple. La fonction qui est en cours d’exécution **s’achève** immédiatement, et l’objet cité dans l’instruction `return` est retourné à l’appelant, qui peut utiliser cette valeur comme n’importe quelle expression.

35.1.4 Le singleton None

Le terme même de fonction, si vous vous rappelez vos souvenirs de mathématiques, suggère qu’on calcule un résultat à partir de valeurs d’entrée. Dans la pratique il est assez rare qu’on définisse une fonction qui ne retourne rien.

En fait **toutes** les fonctions retournent quelque chose. Lorsque le programmeur n’a pas prévu d’instruction `return`, python retourne un objet spécial, baptisé `None`. Voici par exemple ce qu’on obtient si on essaie d’afficher la valeur de retour de notre première fonction, qui, on le rappelle, ne retourne rien :

```
# ce premier appel provoque l'impression d'une ligne
retour = affiche_carre(15)
```

```
le carre de 15 vaut 225
```

```
# voyons ce qu'a retourné la fonction affiche_carre
print('retour =', retour)
```

```
retour = None
```

L’objet `None` est un singleton prédéfini par python, un peu comme `True` et `False`. Ce n’est pas par contre une valeur booléenne, nous aurons l’occasion d’en reparler.

35.1.5 Un exemple un peu plus réaliste

Pour illustrer l’utilisation de `return` sur un exemple plus utile, voyons le code suivant :

```
def premier(n):
    """
    Retourne un booléen selon que n est premier ou non
    Retourne None pour les entrées négatives ou nulles
    """
    # retourne None pour les entrées non valides
    if n <= 0:
        return
    # traiter le cas singulier
    elif n == 1:
        return False
    # chercher un diviseur dans [2..n-1]
    # bien sûr on pourrait s'arrêter à la racine carrée de n
    # mais ce n'est pas notre sujet
    else:
        for i in range(2, n):
            if n % i == 0:
                # on a trouvé un diviseur,
                # on peut sortir de la fonction
                return False
    # à ce stade, le nombre est bien premier
    return True
```

Cette fonction teste si un entier est premier ou non ; il s'agit naturellement d'une version d'école, il existe d'autres méthodes beaucoup plus adaptées à cette tâche. On peut toutefois vérifier que cette version est fonctionnelle pour de petits entiers comme suit. On rappelle que 1 n'est pas considéré comme un nombre premier :

```
for test in [-2, 1, 2, 4, 19, 35]:
    print(f"premier({test:2d}) = {premier(test)}")
```

```
premier(-2) = None
premier( 1) = False
premier( 2) = True
premier( 4) = False
premier(19) = True
premier(35) = False
```

return sans valeur

Pour les besoins de cette discussion, nous avons choisi de retourner `None` pour les entiers négatifs ou nuls, une manière comme une autre de signaler que la valeur en entrée n'est pas valide.

Ceci n'est pas forcément une bonne pratique, mais elle nous permet ici d'illustrer que dans le cas où on ne mentionne pas de valeur de retour, python retourne `None`.

return interrompt la fonction

Comme on peut s'en convaincre en instrumentant le code - ce que vous pouvez faire à titre d'exercice en ajoutant des fonctions `print` - dans le cas d'un nombre qui n'est pas premier la boucle `for` ne va pas jusqu'à son terme.

On aurait pu d'ailleurs tirer profit de cette propriété pour écrire la fonction de manière légèrement différente comme ceci :

```
def premier_sans_else(n):
    """
    Retourne un booléen selon que n est premier ou non
    Retourne None pour les entrées négatives ou nulles
    """
    # retourne None pour les entrées non valides
    if n <= 0:
        return
    # traiter le cas singulier
    if n == 1:
        return False
    # par rapport à la première version, on a supprimé
    # la clause else: qui est inutile
    for i in range(2, n):
        if n % i == 0:
            # on a trouvé un diviseur
            return False
    # a ce stade c'est que le nombre est bien premier
    return True
```

C'est une question de style et de goût. En tous cas, les deux versions sont tout à fait équivalentes, comme on le voit ici :

```
for test in [-2, 2, 4, 19, 35]:
    print(f"pour n ={test:2d} premier → {premier(test)}, "
          f"premier_sans_else → {premier_sans_else(test)}")
```

```
pour n =-2 premier → None, premier_sans_else → None
pour n = 2 premier → True, premier_sans_else → True
pour n = 4 premier → False, premier_sans_else → False
```

```
pour n =19 premier → True, premier_sans_else → True  
pour n =35 premier → False, premier_sans_else → False
```

Digression sur les chaines

Vous remarquerez dans cette dernière cellule, si vous regardez bien le paramètre de `print`, qu'on peut accoler deux chaines (ici deux *f-strings*) sans même les ajouter; un petit détail pour éviter d'alourdir le code :

```
# quand deux chaines apparaissent immédiatement  
# l'une après l'autre sans opérateur, elles sont concaténées  
"abc" "def"
```

```
'abcdef'
```


36.1 Exercice - niveau basique

```
# charger l'exercice
from corrections.exo_label import exo_label
```

Vous devez écrire une fonction qui prend deux arguments

- une chaîne qui désigne le prénom d'un élève
- un entier qui indique la note obtenue

et qui retourne une chaîne, selon que la note est

- $note < 10$
- $10 \leq note < 16$
- $16 \leq note$

comme on le voit sur les exemples :

```
exo_label.example()
```

```
# à vous de jouer
def label(prenom, note):
    "votre code"
```

```
# pour corriger
exo_label.correction(label)
```


37.1 Exercice - niveau basique

37.1.1 Slicing

Commençons par créer une chaîne de caractères. Ne vous inquiétez pas si vous ne comprenez pas encore le code d'initialisation utilisé ci-dessous.

Pour les plus curieux, l'instruction `import` permet de charger dans votre programme une boîte à outils que l'on appelle un module. Python vient avec de nombreux modules qui forment la librairie standard. Le plus difficile avec les modules de la librairie standard est de savoir qu'ils existent. En effet, il y en a un grand nombre et bien souvent il existe un module pour faire ce que vous souhaitez.

Ici en particulier nous utilisons le module `string`.

```
import string
chaine = string.ascii_lowercase
print(chaine)
```

```
abcdefghijklmnopqrstuvwxyz
```

Pour chacune des sous-chaînes ci-dessous, écrire une expression de slicing sur `chaine` qui renvoie la sous-chaîne. La cellule de code doit retourner `True`

Par exemple pour obtenir « `def` » :

```
chaine[3:6] == "def"
```

```
True
```

1. Écrivez une slice pour obtenir « `vwX` » (n'hésitez pas à utiliser les indices négatifs)

```
chaine[ <votre_code> ] == "vwX"
```

2. Une slice pour obtenir « `wxyz` » (avec une seule constante)

```
chaine[ <votre_code> ] == "wxyz"
```

3. Une slice pour obtenir « dfhjlnprtvxz » (avec deux constantes)

```
chaine[ <votre_code> ] == "dfhjlnprtvxz"
```

4. Une slice pour obtenir « xurolifc » (avec deux constantes)

```
chaine[ <votre_code> ] == "xurolifc"
```

37.2 Exercice - niveau intermédiaire

37.2.1 Longueur

```
# il vous faut évaluer cette cellule magique
# pour charger l'exercice qui suit
# et autoévaluer votre réponse
from corrections.exo_inconnue import exo_inconnue
```

On vous donne une chaîne `composite` dont on sait qu'elle a été calculée à partir de deux chaînes `inconnue` et `connue` comme ceci :

```
composite = connue + inconnue + connue
```

On vous donne également la chaîne `connue`. Imaginez par exemple que vous avez (ce ne sont pas les vraies valeurs) :

```
connue = 0bf1
composite = 0bf1a9730e150bf1
```

alors dans ce cas `inconnue` vaut `a9730e15`

L'exercice consiste à écrire une fonction qui retourne la valeur de `inconnue` à partir de celles de `composite` et `connue`.

Vous pouvez utiliser du *slicing*, et la fonction `len()`, qui retourne la longueur d'une chaîne :

```
len('abcd')
```

```
4
```

```
# à vous de jouer
def inconnue(composite, connue):
    return composite[len(connue):-len(connue)]
```

Une fois votre code évalué, vous pouvez évaluer la cellule suivante pour vérifier votre résultat.

```
# correction
exo_inconnue.correction(inconnue)
```

Lorsque vous évaluez cette cellule, la correction vous montre

- dans la première colonne l'appel qui est fait à votre fonction,
- dans la seconde colonne la valeur attendue pour `composite`

— dans la troisième colonne que votre code a réellement calculé.

Si toutes les lignes sont **en vert** c'est que vous avez réussi cet exercice.

Vous pouvez essayer autant de fois que vous voulez, mais il vous faut alors à chaque itération :

- évaluer votre cellule-réponse (là où vous définissez la fonction `inconnue`)
- et ensuite évaluer la cellule correction pour la mettre à jour.

38.1 Exercice - niveau basique

```
from corrections.exo_laccess import exo_laccess
```

Vous devez écrire une fonction `laccess` qui prend en argument une liste, et qui retourne :

- `None` si la liste est vide
- sinon le dernier élément de la liste si elle est de taille paire
- et sinon l'élément du milieu.

```
exo_laccess.example()
```

| Appel | Résultat Attendu |
|--|-------------------|
| <code>laccess([])</code> | <code>None</code> |
| <code>laccess([1])</code> | <code>1</code> |
| <code>laccess(["spam", 2])</code> | <code>2</code> |
| <code>laccess(["spam", 2, "bacon"])</code> | <code>2</code> |

```
# écrivez votre code ici
def laccess(liste):
    return "votre code"
```

```
# pour le corriger
exo_laccess.correction(laccess)
```

Une fois que votre code fonctionne, vous pouvez regarder si par hasard il marcherait aussi avec des chaînes :

```
from corrections.exo_laccess import exo_laccess_strings
```

```
exo_laccess_strings.correction(laccess)
```


Instruction `if` et fonction `def`

39.1 Exercice - niveau basique

39.1.1 Fonction de divisibilité

```
# chargement de l'exercice
from corrections.exo_divisible import exo_divisible
```

L'exercice consiste à écrire une fonction baptisée `divisible` qui retourne une valeur booléenne, qui indique si un des deux arguments est divisible par l'autre.

Vous pouvez supposer les entrées `a` et `b` entiers et non nuls, mais pas forcément positifs.

```
def divisible(a, b):
    "<votre_code>"
```

Vous pouvez à présent tester votre code en évaluant ceci, qui écrira un message d'erreur si un des jeux de test ne donne pas le résultat attendu.

```
# tester votre code
exo_divisible.correction(divisible)
```

39.2 Exercice - niveau basique

```
# chargement de l'exercice
from corrections.exo_morceaux import exo_morceaux
```

On veut définir en python une fonction qui est définie par morceaux :

$$f : x \longrightarrow \begin{cases} -x - 5 & \text{si } x \leq -5 \\ 0 & \text{si } x \in [-5, 5] \\ \frac{1}{5}x - 1 & \text{si } x \geq 5 \end{cases}$$

```
# donc par exemple
exo_morceaux.example()
```

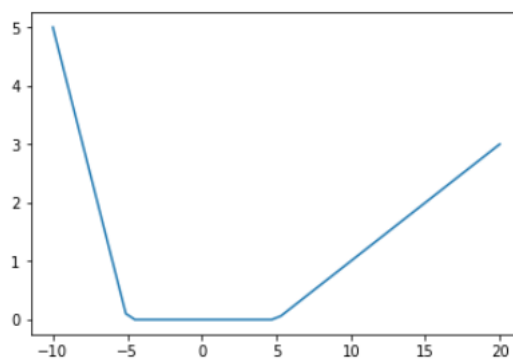
```
# à vous de jouer

def morceaux(x):
    return 0 # "votre code"
```

```
# pour corriger votre code
exo_morceaux.correction(morceaux)
```

L'exercice est maintenant terminé, mais nous allons voir ensemble maintenant comment vous pourriez visualiser votre fonction.

Voici ce qui est attendu comme courbe pour `morceaux` (image fixe) :



En partant de votre code, vous pouvez produire votre propre courbe en utilisant `numpy` et `matplotlib` comme ceci :

```
# on importe les librairies
import numpy as np
import matplotlib.pyplot as plt
```

```
# un échantillon des X entre -10 et 20
X = np.linspace(-10, 20)

# et les Y correspondants
Y = np.vectorize(morceaux)(X)
```

```
# on n'a plus qu'à dessiner
plt.plot(X, Y)
plt.show()
```

40.1 Exercice - niveau basique

40.1.1 Liste des valeurs d'une fonction

```
# Pour charger l'exercice
from corrections.exo_liste_P import exo_liste_P
```

On se donne une fonction polynomiale

$$P(x) = 2x^2 - 3x - 2$$

On vous demande d'écrire une fonction `liste_P` qui prend en argument une liste de nombres réels `x` et qui retourne la liste des valeurs $P(x)$.

```
# voici un exemple de ce qui est attendu
exo_liste_P.example()
```

| Appel | Résultat Attendu |
|---------------------------------------|---------------------------------|
| <code>liste_P([0, 1, 2, 3, 4])</code> | <code>[-2, -3, 0, 7, 18]</code> |

Écrivez votre code dans la cellule suivante (On vous suggère d'écrire une fonction `P` qui implémente le polynôme mais ça n'est pas strictement indispensable, seul le résultat de `"liste_P"` compte) :

```
def P(x):
    "<votre code>"

def liste_P(liste_x):
    "votre code"
```

Et vous pouvez le vérifier en évaluant cette cellule :

```
# pour vérifier votre code
exo_liste_P.correction(liste_P)
```

40.2 Récréation

Si vous avez correctement implémenté la fonction `liste_P` telle que demandé dans le premier exercice, vous pouvez visualiser le polynôme P en utilisant `matplotlib` avec le code suivant :

```
# on importe les librairies
import numpy as np
import matplotlib.pyplot as plt
```

```
# un échantillon des X entre -10 et 10
X = np.linspace(-10, 10)

# et les Y correspondants
Y = liste_P(X)
```

```
# on n'a plus qu'à dessiner
plt.plot(X, Y)
plt.show()
```

41.1 Exercice - niveau intermédiaire

41.1.1 Mise au carré

```
# chargement de l'exercice
from corrections.exo_carre import exo_carre
```

On vous demande à présent d'écrire une fonction dans le même esprit que ci-dessus. Cette fois, chaque ligne contient, séparés par des point-virgules, une liste d'entiers, et on veut obtenir une nouvelle chaîne avec les carrés de ces entiers, séparés par des deux-points.

À nouveau les lignes peuvent être remplies de manière approximative, avec des espaces, des tabulations, ou même des points-virgules en trop, que ce soit au début, à la fin, ou au milieu d'une ligne.

```
# exemples
exo_carre.example()
```

| Appel | Résultat Attendu |
|-----------------------------|---------------------|
| carre("1;2;3") | "1 :4 :9" |
| carre(" 2; 5;6;") | "4 :25 :36" |
| carre(" ; 12; -23;t60; 1t") | "144 :529 :3600 :1" |
| carre(" ; -12;; -23; 1;t") | 144 :529 :1" |

```
<table style="font-family :monospace;font-size :small;"><tr style="font-family :monospace;font-size :medium;"><th style="text-align : center;">Appel</th><th style="text-align : center;">Résultat Attendu</th></tr><tr><td>carre("1;2;3")</td><td><pre> "1 :4 :9"</pre></td></tr><tr><td>carre(" 2; 5;6;")</td><td><pre>  "4 :25 :36"</pre></td></tr><tr><td>carre(" ; 12; -23;t60; 1t")</td><td><pre>  "144 :529 :3600 :1"</pre></td></tr><tr><td>carre(" ; -12;; -23; 1;t")</td><td><pre> "144 :529 :1"</pre></td></tr></table>
```

```
# écrivez votre code ici
def carre(ligne) :
    "<votre_code>"
```

```
# pour corriger  
exo_carre.correction(carre)
```

42.1 Complément - niveau basique

Voici quelques utilisations habituelles du type fichier en python

42.1.1 Avec un *context manager*

Nous avons vu dans la vidéo les mécanismes de base sur les fichiers. Nous avons vu notamment qu'il est important de bien fermer un fichier après usage. On a vu aussi qu'il est recommandé de **toujours** utiliser l'instruction `with` et de contrôler son encodage. Il est donc recommandé de faire :

```
# avec un 'with' on garantit la fermeture du fichier
with open("foo.txt", "w", encoding='utf-8') as sortie:
    for i in range(2):
        sortie.write(f"{i}\n")
```

42.1.2 Les modes d'ouverture

Les modes d'ouverture les plus utilisés sont

- 'r' (la chaîne contenant l'unique caractère `r`) pour ouvrir un fichier en lecture seulement ;
- 'w' en écriture seulement ; le contenu précédent du fichier, s'il existait, est perdu ;
- 'a' en écriture seulement, mais pour ajouter du contenu en fin de fichier.

Voici par exemple comment on pourrait ajouter deux lignes de texte dans le fichier `foo.txt` qui contient, à ce stade du notebook, 2 entiers :

```
# on ouvre le fichier en mode 'a' comme append (= ajouter)
with open("foo.txt", "a", encoding='utf-8') as sortie:
    for i in range(100, 102):
        sortie.write("{}\n".format(i))
```

```
# maintenant on regarde ce que contient le fichier
with open("foo.txt", encoding='utf-8') as entree: # remarquez que sans 'mode', on
    ouvre en lecture seule
```

```
for line in entree:
    # line contient déjà un newline
    print(line, end='')
```

```
0
1
100
101
```

Il existe de nombreuses variantes au mode d'ouverture, pour par exemple :

- ouvrir le fichier en lecture *et* en écriture (mode +),
- ouvrir le fichier en mode binaire (mode b).

Ces variantes sont décrites dans *la section sur la fonction built-in ‘open’* <<https://docs.python.org/3/library/functions.html#open>>‘__ dans la documentation python.

42.2 Complément - niveau intermédiaire

42.2.1 Un fichier est un itérateur

Nous reparlerons des notions d'itérable et d'itérateur dans les semaines suivantes. Pour l'instant, on peut dire qu'un fichier - qui donc **est itérable** puisqu'on peut le lire par une boucle `for` - est aussi **son propre itérateur**. Cela implique que l'on ne peut le parcourir qu'une fois dans une boucle `for`. Pour le reparcourir, il faut le fermer et l'ouvrir de nouveau.

```
# un fichier est son propre itérateur
```

```
with open("foo.txt", encoding='utf-8') as entree:
    print(entree.__iter__() is entree)
```

```
True
```

Par conséquent, écrire deux boucles `for` imbriquées sur **le même objet fichier** ne **fonctionnerait pas** comme on pourrait s'y attendre.

```
# Si on essaie d'écrire deux boucles imbriquées
# sur le même objet fichier, le résultat est inattendu
with open("foo.txt", encoding='utf-8') as entree:
    for l1 in entree:
        # on enlève les fins de ligne
        l1 = l1.strip()
        for l2 in entree:
            # on enlève les fins de ligne
            l2 = l2.strip()
            print(l1, "x", l2)
```

```
0 x 1
0 x 100
0 x 101
```


42.3 Complément - niveau avancé

42.3.1 Autres méthodes

Vous pouvez également accéder à des fonctions de beaucoup plus bas niveau, notamment celle fournies directement par le système d'exploitation ; nous allons en décrire deux parmi les plus utiles.

Digression - `repr()`

Comme nous allons utiliser maintenant des outils d'assez bas niveau pour lire du texte, aussi pour examiner ce texte nous allons utiliser la fonction `repr()`, et voici pourquoi :

```
# construisons à la main une chaîne qui contient deux lignes
lines = "abc" + "\n" + "def" + "\n"
```

```
# si on l'imprime on voit bien les newline
# d'ailleurs on sait qu'il n'est pas utile
# d'ajouter un newline à la fin
print(lines, end="")
```

```
abc
def
```

```
# vérifions que repr() nous permet de bien
# voir le contenu de cette chaîne
print(repr(lines))
```

```
'abcndefn'
```

Lire un contenu - bas niveau

Revenons aux fichiers ; la méthode `read()` permet de lire dans le fichier un buffer d'une certaine taille :

```
# read() retourne TOUT le contenu
# ne pas utiliser avec de très gros fichier bien sûr

# une autre façon de montrer tout le contenu du fichier
with open("foo.txt", encoding='utf-8') as entree:
    full_contents = entree.read()
    print("Contenu complet\n{}".format(full_contents), end="")
```

```
Contenu complet
0
1
100
101
```

```
# lire dans le fichier deux blocs de 4 caractères
with open("foo.txt", encoding='utf-8') as entree:
    for bloc in range(2):
        print("Bloc {} >>{}<<".format(bloc, repr(entree.read(4))))
```

```
Bloc 0 >>'0n1n'<<
```

```
Bloc 1 >>'100n'<<
```

On voit donc que chaque bloc contient bien 4 caractères en comptant les sauts de ligne

| bloc # | contenu |
|--------|---|
| 0 | un 0, un <i>newline</i> , un 1, un <i>newline</i> |
| 1 | un 1, deux 0, un <i>newline</i> |

La méthode `flush`

Les entrées-sortie sur fichier sont bien souvent *bufferisées* par le système d'exploitation. Cela signifie qu'un appel à `write` ne provoque pas forcément une écriture immédiate, car pour des raisons de performance on attend d'avoir suffisamment de matière avant d'écrire sur le disque.

Il y a des cas où ce comportement peut s'avérer gênant, et où on a besoin d'écrire immédiatement (et donc de vider le *buffer*), et c'est le propos de la méthode `flush()`.

42.3.2 Fichiers textuels et fichiers binaires

De la même façon que le langage propose les deux types `str` et `bytes`, il est possible d'ouvrir un fichier en mode *textuel* ou en mode *binaire*.

Les fichiers que nous avons vus jusqu'ici étaient ouverts en mode *textuel* (c'est le défaut), et c'est pourquoi quand nous avons interagi avec eux avec des objets de type `str` :

```
# un fichier ouvert en mode textuel nous donne des str
with open('foo.txt', encoding='utf-8') as input:
    for line in input:
        print("on a lu un objet de type", type(line))
```

```
on a lu un objet de type <class 'str'>
on a lu un objet de type <class 'str'>
on a lu un objet de type <class 'str'>
on a lu un objet de type <class 'str'>
```

Lorsque ce n'est pas le comportement souhaité, on peut

- ouvrir le fichier en mode *binaire* - pour cela on ajoute le caractère `b` au mode d'ouverture
- et on peut alors interagir avec le fichier avec des objets de type `bytes`

Pour illustrer ce trait, nous allons : 1. créer un fichier en mode texte, et y insérer du texte en UTF-8 1. relire le fichier en mode binaire, et retrouver le codage des différents caractères.

```
# phase 1 : on écrit un fichier avec du texte en UTF-8
# on ouvre le donc le fichier en mode texte
# en toute rigueur il faut préciser l'encodage,
# si on ne le fait pas il sera déterminé
# à partir de vos réglages système
with open('strbytes', 'w', encoding='utf-8') as output:
    output.write("déjà l'été\n")
```

```
# phase 2: on rouvre le fichier en mode binaire
with open('strbytes', 'rb') as rawinput:
    # on relit tout le contenu
    octets = rawinput.read()
    # qui est de type bytes
    print("on a lu un objet de type", type(octets))
    # si on regarde chaque octet un par un
    for i, octet in enumerate(octets):
        print("{} → {} [{}]" .format(i, repr(chr(octet)), hex(octet)))
```

```

on a lu un objet de type <class 'bytes'>
0 → 'd' [0x64]
1 → 'Ã' [0xc3]
2 → '©' [0xa9]
3 → 'j' [0x6a]
4 → 'Ã' [0xc3]
5 → 'xa0' [0xa0]
6 → ' ' [0x20]
7 → 'l' [0x6c]
8 → '"' [0x27]
9 → 'Ã' [0xc3]
10 → '©' [0xa9]
11 → 't' [0x74]
12 → 'Ã' [0xc3]
13 → '©' [0xa9]
14 → 'r' [0xd]
15 → 'n' [0xa]

```

Vous retrouvez ainsi le fait que l’unique caractère unicode “é”, a été encodé par UTF-8 sous la forme de deux octets de code hexadécimal 0xc3 et 0xa9.

Vous pouvez également consulter ce site qui visualise l’encodage UTF-8, avec notre séquence d’entrée

<https://mothereff.in/utf-8#d%C3%A9%C3%A0%20l%27%C3%A9t%C3%A9%0A>

```

# on peut comparer le nombre d'octets et le nombre de caractères
with open('strbytes', encoding='utf-8') as textfile:
    print("en mode texte, {} caractères".format(len(textfile.read())))
with open('strbytes', 'rb') as binfile:
    print("en mode binaire, {} octets".format(len(binfile.read())))

```

```

en mode texte, 11 caractères
en mode binaire, 16 octets

```

Ce qui correspond au fait que nos 4 caractères non-ASCII (3 é et 1 à) sont tous encodés par UTF-8 comme 2 octets, comme vous pouvez vous en assurer *ici* pour “é” <<https://mothereff.in/utf-8#%C3%A9>>‘__ et *là* pour “à” <<https://mothereff.in/utf-8#%C3%A0>>‘__.

42.3.3 Pour en savoir plus

Pour une description exhaustive vous pouvez vous reporter à

- au *glossaire sur la notion de “object file”* <<https://docs.python.org/3/glossary.html#term-file-object>>‘__,
- et aussi et surtout au *module “io”* <<https://docs.python.org/3/library/io.html#module-io>>‘__ qui décrit plus en détails les fonctionnalités disponibles.

43.1 Complément - niveau basique

Outre les objets fichiers créés avec la fonction `open`, comme on l'a vu dans la vidéo, et qui servent à lire et écrire à un endroit précis, une application a besoin d'un minimum d'utilitaires pour **parcourir l'arborescence de répertoires et fichiers**, c'est notre propos dans ce complément.

43.1.1 Le module `os.path` (obsolète)

Avant la version python-3.4, la librairie standard offrait une conjonction d'outils pour ce type de fonctionnalités :

- le module `os.path`, pour faire des calculs sur les chemins et noms de fichiers [doc](#),
- le module `os` pour certaines fonctions complémentaires comme renommer ou détruire un fichier [doc](#),
- et enfin le module `glob` pour la recherche de fichiers, par exemple pour trouver tous les fichiers en `*.txt` [doc](#).

Cet ensemble un peu disparate a été remplacé par une librairie unique `pathlib`, qui fournit toutes ces fonctionnalités sous un interface unique et moderne, que nous recommandons évidemment d'utiliser pour du nouveau code.

Avant d'aborder `pathlib`, voici un très bref aperçu de ces trois anciens modules, pour le cas - assez probable - où vous les rencontreriez dans du code existant; tous les noms qui suivent correspondent à des **fonctions** - par opposition à `pathlib` qui comme nous allons le voir offre une interface orientée objet :

- `os.path.join` ajoute `'/'` ou `'\"` entre deux morceaux de chemin, selon l'OS
- `os.path.basename` trouve le nom de fichier dans un chemin
- `os.path.dirname` trouve le nom du directory dans un chemin
- `os.path.abspath` calcule un chemin absolu, c'est-à-dire à partir de la racine du filesystem
- `os.path.exists` pour savoir si un chemin existe ou pas (fichier ou répertoire)
- `os.path.isfile` (et `isdir`) pour savoir si un chemin est un fichier (et un répertoire)
- `os.path.getsize` pour obtenir la taille du fichier
- `os.path.getatime` et aussi `getmtime` et `getctime` pour obtenir les dates de création/modification d'un fichier
- `os.remove` (ou son ancien nom `os.unlink`), qui permet de supprimer un fichier

- `os.rmdir` pour supprimer un répertoire (mais qui doit être vide)
- `os.removedirs` pour supprimer tout un répertoire avec son contenu, récursivement si nécessaire
- `os.rename` pour renommer un fichier
- `glob.glob` comme dans par exemple `glob.glob("*.txt")`

43.1.2 Le module `pathlib`

Orienté Objet

Comme on l'a mentionné `pathlib` offre une interface orientée objet ; mais qu'est-ce que ça veut dire au juste ?

Ceci nous donne un prétexte pour une première application pratique des notions de module (que nous avons introduits en fin de semaine 2) et de classe (que nous allons voir en fin de semaine).

De même que le langage nous propose les types *builtin* `int` et `str`, le module `pathlib` nous expose **un type** (on dira plutôt **une classe**) qui s'appelle `Path`, que nous allons importer comme ceci :

```
from pathlib import Path
```

Nous allons faire tourner un petit scénario qui va créer un fichier :

```
# le nom de notre fichier jouet
nom = 'fichier-temoin'
```

Pour commencer, nous allons vérifier si le fichier en question existe.

Pour ça nous créons un **objet** qui est une **instance** de la classe `Path`, comme ceci :

```
# on crée un objet de la classe Path, associé au nom de fichier
path = Path(nom)
```

Vous remarquez que c'est consistant avec par exemple :

```
# transformer un float en int
i = int(3.5)
```

en ce sens que le type (`int` ou `Path`) se comporte comme une usine pour créer des objets du type en question.

Quoi qu'il en soit, cet objet `path` offre un certain nombre de méthodes ; pour les voir puisque nous sommes dans un notebook, je vous invite dans la cellule suivante à utiliser l'aide en ligne en appuyant sur la touche 'Tabulation' après avoir ajouté un `.` comme si vous alliez envoyer une méthode à cet objet

```
path.[taper la touche TAB]
```

et le notebook vous montrera la liste des méthodes disponibles.

```
# ajouter un . et utilisez la touche <Tabulation>
path
```

```
WindowsPath('fichier-temoin')
```

Ainsi par exemple on peut savoir si le fichier existe avec la méthode `exists()`

```
# au départ le fichier n'existe pas
path.exists()
```

```
False
```

```
# si j'écris dedans je le crée
with open(nom, 'w', encoding='utf-8') as output:
    output.write('0123456789\n')
```

```
# et maintenant il existe
path.exists()
```

```
True
```

métadonnées

Voici quelques exemples qui montrent comment accéder aux métadonnées de ce fichier :

```
# cette méthode retourne (en un seul appel système) les métadonnées agrégées
path.stat()
```

```
os.stat_result(st_mode=33206, st_ino=29836347531488470, st_dev=1319427725, st_
↪nlink=1, st_uid=0, st_gid=0, st_size=12, st_atime=1510845651, st_
↪mtime=1510845651, st_ctime=1510845651)
```

Pour ceux que ça intéresse, l'objet retourné par cette méthode `stat` est un `namedtuple`, que l'on va voir très bientôt.

On accède aux différentes informations comme ceci :

```
# la taille du fichier en octets est de 11
# car il faut compter un caractère "newline" en fin de ligne
path.stat().st_size
```

```
12
```

```
# la date de dernière modification, sous forme d'un entier
# c'est le nombre de secondes depuis le 1er Janvier 1970
mtime = path.stat().st_mtime
mtime
```

```
1510845651.7996907
```

```
# que je peux rendre lisible comme ceci
# en anticipant sur le module datetime
from datetime import datetime
mtime_datetime = datetime.fromtimestamp(mtime)
mtime_datetime
```

```
datetime.datetime(2017, 11, 16, 16, 20, 51, 799691)
```

```
# ou encore, si je formate pour n'obtenir que
# l'heure et la minute
f"{mtime_datetime:%H:%M}"
```

```
'16:20'
```

Détruire un fichier

```
# je peux maintenant détruire le fichier
path.unlink()
```

```
# ou encore mieux, si je veux détruire
# seulement dans le cas où il existe je peux aussi faire
```

```
try:
    path.unlink()
except FileNotFoundError:
    print("no need to remove")
```

```
no need to remove
```

```
# et maintenant il n'existe plus
path.exists()
```

```
False
```

```
# je peux aussi retrouver le nom du fichier comme ceci
# attention ce n'est pas une méthode mais un attribut
# c'est pourquoi il n'y a pas de parenthèses
path.name
```

```
'fichier-temoin'
```

Recherche de fichiers

Maintenant je voudrais connaître la liste des fichiers de nom *.json dans le directory data.

La méthode la plus naturelle consiste à créer une instance de Path associée au directory lui-même :

```
dirpath = Path('./data/')
```

Sur cet objet la méthode glob nous retourne un itérable qui contient ce qu'on veut :

```
# tous les fichiers *.json dans le répertoire data/
for json in dirpath.glob("*.json"):
    print(json)
```

Documentation complète

Voyez la documentation complète [ici](#)

43.2 Complément - niveau avancé

Pour ceux qui sont déjà familiers avec les classes, j'en profite pour vous faire remarquer le type de notre objet path

```
type(path)
```

```
pathlib.WindowsPath
```

qui n'est pas Path, mais en fait une sous-classe de Path qui est - sur la plateforme du MOOC au moins, qui fonctionne sous linux - un objet de type PosixPath, qui est une sous-classe de Path, comme vous pouvez le voir :

```
from pathlib import PosixPath
issubclass(PosixPath, Path)
```

```
True
```


Ce qui fait que mécaniquement, `path` est bien une instance de `Path`

```
isinstance(path, Path)
```

```
True
```

ce qui est heureux puisqu'on avait utilisé `Path()` pour construire l'objet `path` au départ :)

Formats de fichiers : JSON et autres

44.1 Compléments - niveau basique

Voici quelques mots sur des outils python fournis dans la librairie standard, et qui permettent de lire ou écrire des données dans des fichiers.

44.1.1 Le problème

Les données dans un programme python sont stockés en mémoire (la RAM), sous une forme propice aux calculs. Par exemple un petit entier est fréquemment stocké en binaire dans un mot de 64 bits, qui est prêt à être soumis au processeur pour faire une opération arithmétique.

Ce format ne se prête pas forcément toujours à être transposé tel quel lorsqu'on doit écrire des données sur un support plus pérenne, comme un disque dur, ou encore sur un réseau pour transmission distante - ces deux supports étant à ce point de vue très voisins.

Ainsi par exemple il pourra être plus commode d'écrire notre entier sur disque, ou de le transmettre à un programme distant, sous une forme décimale qui sera plus lisible, sachant que par ailleurs toutes les machines ne codent pas un entier de la même façon.

Il convient donc de faire de la traduction dans les deux sens entre représentations d'une part en mémoire, et d'autre part sur disque ou sur réseau (à nouveau, on utilise en général les mêmes formats pour ces deux usages).

44.1.2 Le format JSON

Le format sans aucun doute le plus populaire à l'heure actuelle est le **format JSON** pour *JavaScript Object Notation*.

Sans trop nous attarder nous dirons que JSON est un encodage - en anglais **marshalling** - qui se prête bien à la plupart des types de base qu'on trouve dans les langages modernes comme python, ruby ou JavaScript.

La librairie standard python contient le **module json** que nous illustrons très rapidement ici :

```
import json

# En partant d'une donnée construite à partir de types de base
data = [
    # des types qui ne posent pas de problème
```

```

[1, 2, 'a', [3.23, 4.32], {'eric': 32, 'jean': 43}],
# un tuple
(1, 2, 3),
]

# sauver ceci dans un fichier
with open("s1.json", "w", encoding='utf-8') as json_output:
    json.dump(data, json_output)

# et relire le résultat
with open("s1.json", encoding='utf-8') as json_input:
    data2 = json.load(json_input)

```

Limitations de json

Certains types de base ne sont pas supportés par le format JSON (car ils ne sont pas natifs en JavaScript), c'est le cas notamment de :

- tuple, qui se fait encoder comme une liste;
- complex, set et frozenset, qu'on ne peut pas encoder du tout (sans étendre la librairie).

C'est ce qui explique ce qui suit :

```

# le premier élément de data est intact,
# comme si on avait fait une *deep copy* en fait
print("première partie de data", data[0] == data2[0])

```

```
première partie de data True
```

```

# par contre le `tuple` se fait encoder comme une `list`
print("deuxième partie", "entrée", type(data[1]), "sortie", type(data2[1]))

```

```
deuxième partie entrée <class 'tuple'> sortie <class 'list'>
```

Malgré ces petites limitations, ce format est de plus en plus populaire, notamment parce qu'on peut l'utiliser pour communiquer avec des applications web écrites en JavaScript, et aussi parce qu'il est très léger, et supporté par de nombreux langages.

44.2 Compléments - niveau intermédiaire

44.2.1 Le format csv

Le format `csv` pour *Comma Separated Values*, originaire du monde des tableurs, peut rendre service à l'occasion, il est proposé dans le module `'csv'` <<https://docs.python.org/3/library/csv.html>>__.

44.2.2 Le format pickle

Le format `pickle` remplit une fonctionnalité très voisine de JSON, mais est spécifique à python. C'est pourquoi, malgré des limites un peu moins sévères, son usage tend à rester plutôt marginal pour l'échange de données, on lui préfère en général le format JSON.

Par contre, pour la sauvegarde locale d'objets python (pour, par exemple, faire des points de reprises d'un programme), il est très utile. Il est implémenté dans le module `'pickle'` <<https://docs.python.org/3/library/pickle.html>>__.

44.2.3 Le format XML

Vous avez aussi très probablement entendu parler de XML, qui est un format assez populaire également.

Cela dit, la puissance, et donc le coût, de XML et JSON ne sont pas du tout comparables, XML étant beaucoup plus flexible mais au prix d'une complexité de mise en œuvre très supérieure.

Il existe plusieurs souches différentes de bibliothèques prenant en charge le format XML, [qui sont introduites ici](#).

44.2.4 Pour en savoir plus

Voyez la page sur [les formats de fichiers](#) dans la documentation python.

45.1 Complément - niveau avancé

Dans ce complément, nous allons voir comment un programme python interagit avec ce qu'il est convenu d'appeler le système d'entrées-sorties standard du système d'exploitation.

45.1.1 Introduction

Dans un ordinateur, le système d'exploitation (Windows, Linux, ou MacOS) est un logiciel (*kernel*) qui a l'exclusivité pour interagir physiquement avec le matériel (CPU, mémoire, disques, périphériques, etc.); il offre aux programmes utilisateur (*userspace*) des abstractions pour interagir avec ce matériel.

La notion de fichier, telle qu'on l'a vue dans la vidéo, correspond à une de ces abstractions; elle repose principalement sur les 4 opérations élémentaires

- open
- close
- read
- write

Parmi les autres conventions d'interaction entre le système (pour être précis : le **shell**) et une application, il y a les notions de

- entrée standard (*standard input*, en abrégé `stdin`)
- sortie standard (*standard output*, en abrégé `stdout`)
- erreur standard (*standard error*, en abrégé `stderr`)

Ceci est principalement pertinent dans le contexte d'un terminal. L'idée c'est qu'on a envie de pouvoir **rediriger** les entrées-sorties d'un programme sans avoir à le modifier. De la sorte, on peut également *chaîner* des traitements à l'aide de **pipes**, sans avoir besoin de sauver les résultats intermédiaires sur disque.

Ainsi par exemple lorsqu'on écrit

```
$ monprogramme < fichier_entree > fichier_sortie
```

les deux fichiers en question sont ouverts par le *shell*, et passés à `monprogramme` - que celui-ci soit écrit en C, en python ou en Java - sous la forme des fichiers `stdin` et `stdout` respectivement, et donc **déjà ouverts**.

45.1.2 Le module `sys`

L'interpréteur python vous expose ces trois fichiers sous la forme d'attributs du module `sys` :

```
import sys
for channel in (sys.stdin, sys.stdout, sys.stderr):
    print(channel)
```

```
<_io.TextIOWrapper name='<stdin>' mode='r' encoding='cp1252'>
<ipykernel.iostream.OutputStream object at 0x03F929F0>
<ipykernel.iostream.OutputStream object at 0x03FA3F70>
```

Dans le contexte du notebook vous pouvez constater que les deux flux de sortie sont implémentés comme des classes spécifiques à IPython. Si vous exécutez ce code localement dans votre ordinateur vous allez sans doute obtenir quelque chose comme

```
<_io.TextIOWrapper name='<stdin>' mode='r' encoding='UTF-8'>
<_io.TextIOWrapper name='<stdout>' mode='w' encoding='UTF-8'>
<_io.TextIOWrapper name='<stderr>' mode='w' encoding='UTF-8'>
```

On n'a pas extrêmement souvent besoin d'utiliser ces variables en règle générale, mais elles peuvent s'avérer utiles dans des contextes spécifiques.

Par exemple, l'instruction `print` écrit dans `sys.stdout` (c'est-à-dire la sortie standard). Et comme `sys.stdout` est une variable (plus exactement `stdout` est un attribut dans le module référencé par la variable `sys`) et qu'elle référence un objet fichier, on peut lui faire référencer un autre objet fichier et ainsi rediriger depuis notre programme tous les sorties, qui sinon iraient sur le terminal, vers un fichier de notre choix :

```
# ici je fais exprès de ne pas utiliser un `with`
# car très souvent les deux redirections apparaissent
# dans des fonctions différentes
import sys
# on ouvre le fichier destination
autre_stdout = open('ma_sortie.txt', 'w', encoding='utf-8')
# on garde un lien vers le fichier sortie standard
# pour le réinstaller plus tard si besoin.
tmp = sys.stdout
#
print('sur le terminal')
# première redirection
sys.stdout = autre_stdout
#
print('dans le fichier')
# on remet comme c'était au début
sys.stdout = tmp
# et alors pour être propre on n'oublie pas de fermer
autre_stdout.close()
#
print('de nouveau sur le terminal')
```

```
sur le terminal
de nouveau sur le terminal
```



```
# et en effet, dans le fichier on a bien  
with open("ma_sortie.txt", encoding='utf-8') as check:  
    print(check.read())
```

dans le fichier

La construction de tuples

46.1 Complément - niveau intermédiaire

46.1.1 Les tuples et la virgule terminale

Comme on l'a vu dans la vidéo, on peut construire un tuple à deux éléments - un couple - de quatre façons :

```
# sans parenthèse ni virgule terminale
couple1 = 1, 2
# avec parenthèses
couple2 = (1, 2)
# avec virgule terminale
couple3 = 1, 2,
# avec parenthèse et virgule
couple4 = (1, 2,)
```

```
# toutes ces formes sont équivalentes; par exemple
couple1 == couple4
```

```
True
```

Comme on le voit :

- en réalité la **parenthèse est parfois superflue** ; mais il se trouve qu'elle est **largement utilisée** pour améliorer la lisibilité des programmes, sauf dans le cas du *tuple unpacking* ; nous verrons aussi plus bas qu'elle est **parfois nécessaire** selon l'endroit où le tuple apparaît dans le programme ;
- la **dernière virgule est optionnelle** aussi, c'est le cas pour les tuples à au moins 2 éléments - nous verrons plus bas le cas des tuples à un seul élément.

46.1.2 Conseil pour la présentation sur plusieurs lignes

En général d'ailleurs, la forme avec parenthèses et virgule terminale est plus pratique. Considérez par exemple l'initialisation suivante ; on veut créer un tuple qui contient des listes (naturellement un tuple peut contenir n'importe quel objet python), et comme c'est assez long on préfère mettre un élément du tuple par ligne :

```
mon_tuple = ([1, 2, 3],
             [4, 5, 6],
             [7, 8, 9],
             )
```

L'avantage lorsqu'on choisit cette forme (avec parenthèses, et avec virgule terminale), c'est que d'abord il n'est pas nécessaire de mettre un backslash à la fin de chaque ligne ; parce que l'on est à l'intérieur d'une zone parenthésée, l'interpréteur python « sait » que l'instruction n'est pas terminée et va se continuer sur la ligne suivante.

Deuxièmement, si on doit ultérieurement ajouter ou enlever un élément dans le tuple, il suffira d'enlever ou d'ajouter toute une ligne, sans avoir à s'occuper des virgules ; si on avait choisi de ne pas faire figurer la virgule terminale, alors pour ajouter un item dans le tuple après le dernier, il ne faut pas oublier d'ajouter une virgule à la ligne précédente. Cette simplicité se répercute au niveau du gestionnaire de code source, où les différences dans le code sont plus faciles à visualiser.

Signalons enfin que ceci n'est pas propre aux tuples. La virgule terminale est également optionnelle pour les listes, ainsi d'ailleurs que pour tous les types python où cela fait du sens, comme les dictionnaires et les ensembles que nous verrons bientôt. Et dans tous les cas où on opte pour une présentation multi-lignes, il est conseillé de faire figurer une virgule terminale.

46.1.3 Tuples à un élément

Pour revenir à présent sur le cas des tuples à un seul élément, c'est un cas particulier, parmi les 4 syntaxes qu'on a vues ci-dessus, on obtiendrait dans ce cas

```
# ATTENTION: ces deux premières formes ne construisent pas un tuple !
simple1 = 1
simple2 = (1)
# celles-ci par contre construisent bien un tuple
simple3 = 1,
simple4 = (1,)
```

- Il est bien évident que la première forme ne crée pas de tuple ;
- et en fait la seconde non plus, car python lit ceci comme une expression parenthésée, avec seulement un entier.

Et en fait ces deux premières formes créent un entier simple :

```
type(simple2)
```

```
int
```

Les deux autres formes créent par contre toutes les deux un tuple à un élément comme on cherchait à le faire :

```
type(simple3)
```

```
tuple
```

```
simple3 == simple4
```

```
True
```

Pour conclure, disons donc qu'il est conseillé de **toujours mentionner une virgule terminale** lorsqu'on construit des tuples.

46.1.4 Parenthèse parfois obligatoire

Dans certains cas vous vous apercevrez que la parenthèse est obligatoire. Par exemple on peut écrire :

```
x = (1,)
(1,) == x
```

True

Mais si on essaie d'écrire le même test sans les parenthèses :

```
# ceci provoque une SyntaxError
1, == x
```

```
File "<ipython-input-9-219921aa55c4>", line 2
  1, == x
    ^
SyntaxError: invalid syntax
```

Python lève une erreur de syntaxe ; encore une bonne raison pour utiliser les parenthèses.

46.1.5 Addition de tuples

Bien que le type tuple soit immuable, il est tout à fait légal d'additionner deux tuples, et l'addition va produire un **nouveau** tuple.

```
tuple1 = (1, 2,)
tuple2 = (3, 4,)
print('addition', tuple1 + tuple2)
```

```
addition (1, 2, 3, 4)
```

Ainsi on peut également utiliser l'opérateur += avec un tuple qui va créer, comme précédemment, un nouvel objet tuple :

```
tuple1 = (1, 2,)
tuple1 += (3, 4,)
print('apres ajout', tuple1)
```

```
apres ajout (1, 2, 3, 4)
```

46.1.6 Construire des tuples élaborés

Malgré la possibilité de procéder par additions successives, la construction d'un tuple peut s'avérer fastidieuse.

Une astuce utile consiste à penser aux fonctions de conversion, pour construire un tuple à partir de - par exemple - une liste. Ainsi on peut faire par exemple ceci :

```
# on fabrique une liste pas à pas
liste = list(range(10))
liste[9] = 'Inconnu'
del liste [2:5]
liste
```

```
[0, 1, 5, 6, 7, 8, 'Inconnu']
```

```
# on convertit le résultat en tuple
mon_tuple = tuple(liste)
mon_tuple
```

```
(0, 1, 5, 6, 7, 8, 'Inconnu')
```

46.1.7 Digression sur les noms de fonctions prédéfinies

Remarque. Vous avez peut-être observé que nous avons choisi de ne pas appeler notre tuple simplement `tuple`. C'est une bonne pratique en général d'éviter les noms de fonctions prédéfinies par python.

Ces variables en effet sont des variables « comme les autres ». Imaginez qu'on ait en fait deux tuples à construire comme ci-dessus, voici ce qu'on obtiendrait si on n'avait pas pris cette précaution

```
liste = range(10)
# ATTENTION: ceci redéfinit le symbole tuple
tuple = tuple(liste)
tuple
```

```
(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
```

```
# si bien que maintenant on ne peut plus faire ceci
# car à ce point, tuple ne désigne plus le type tuple
# mais l'objet qu'on vient de créer
autre_liste = range(100)
autre_tuple = tuple(autre_liste)
```

```
-----
TypeError                                Traceback (most recent call last)

<ipython-input-15-15ef2d1e6804> in <module>()
      3 # mais l'objet qu'on vient de créer
      4 autre_liste = range(100)
----> 5 autre_tuple = tuple(autre_liste)

TypeError: 'tuple' object is not callable
```

Il y a une erreur parce que nous avons remplacé (ligne 2) la valeur de la variable `tuple`, qui au départ référençait le **type** `tuple` (ou si on préfère le fonction de conversion), par un **objet** `tuple`. Ainsi en ligne 5, lorsqu'on appelle à nouveau `tuple`, on essaie d'exécuter un objet qui n'est pas "appelable" (*not callable* en anglais).

D'un autre côté, l'erreur est relativement facile à trouver dans ce cas. En cherchant toutes les occurrences de `tuple` dans notre propre code on voit assez vite le problème. De plus, je vous rappelle que votre éditeur de texte **doit** faire de la coloration syntaxique, et que toutes les fonctions built-in (dont `tuple` et `list` font partie) sont colorées spécifiquement (par exemple, en violet sous IDLE). En pratique, avec un bon éditeur de texte et un peu d'expérience, cette erreur est très rare.

Sequence unpacking

47.1 Complément - niveau basique

Remarque préliminaire : nous avons vainement cherché une traduction raisonnable pour ce trait du langage, connue en anglais sous le nom de *sequence unpacking* ou encore parfois *tuple unpacking*, aussi pour éviter de créer de la confusion nous avons finalement décidé de conserver le terme anglais à l'identique.

47.1.1 Déjà rencontré

L'affectation dans python peut concerner plusieurs variables à la fois. En fait nous en avons déjà vu un exemple en Semaine 1, avec la fonction `fibonacci` dans laquelle il y avait ce fragment :

```
for i in range(2, n + 1):  
    f2, f1 = f1, f1 + f2
```

Nous allons dans ce complément décortiquer les mécanismes derrière cette phrase qui a probablement excité votre curiosité :)

47.1.2 Un exemple simple

Commençons par un exemple simple à base de tuple. Imaginons qu'on dispose d'un tuple `couple` dont on sait qu'il a deux éléments :

```
couple = (100, 'spam')
```

On souhaite à présent extraire les deux valeurs, et les affecter à deux variables distinctes. Une solution naïve consiste bien sûr à faire simplement :

```
gauche = couple[0]  
droite = couple[1]  
print('gauche', gauche, 'droite', droite)
```

```
gauche 100 droite spam
```

Cela fonctionne naturellement très bien, mais n'est pas très pythonique - comme on dit;) Vous devez toujours garder en tête qu'il est rare en python de manipuler des indices. Dès que vous voyez des indices dans votre code, vous devez vous demander si votre code est pythonique.

On préférera la formulation équivalente suivante :

```
(gauche, droite) = couple
print('gauche', gauche, 'droite', droite)
```

```
gauche 100 droite spam
```

La logique ici consiste à dire, affecter les deux variables de sorte que le tuple (gauche, droite) soit égal à couple. On voit ici la supériorité de cette notion d'unpacking sur la manipulation d'indices : vous avez maintenant des variables qui expriment la nature de l'objet manipulé, votre code devient expressif, c'est-à-dire auto-documenté.

Remarquons que les parenthèses ici sont optionnelles - comme lorsqu'on construit un tuple - et on peut tout aussi bien écrire, et c'est le cas d'usage le plus fréquent d'omission des parenthèses pour le tuple :

```
gauche, droite = couple
print('gauche', gauche, 'droite', droite)
```

```
gauche 100 droite spam
```

47.1.3 Autres types

Cette technique fonctionne aussi bien avec d'autres types. Par exemple je peux utiliser :

- une syntaxe de liste à gauche du =
- une liste comme expression à droite du =

```
# comme ceci
liste = [1, 2, 3]
[gauche, milieu, droit] = liste
print('gauche', gauche, 'milieu', milieu, 'droit', droit)
```

```
gauche 1 milieu 2 droit 3
```

Et on n'est même pas obligés d'avoir le même type à gauche et à droite du signe =, comme ici :

```
# membre droit: une liste
liste = [1, 2, 3]
# membre gauche : un tuple
gauche, milieu, droit = liste
print('gauche', gauche, 'milieu', milieu, 'droit', droit)
```

```
gauche 1 milieu 2 droit 3
```

En réalité, les seules contraintes fixées par python sont que

- le terme à droite du signe = est un *iterable* (tuple, liste, string, etc.),
- le terme à gauche soit écrit comme un tuple ou une liste - notons tout de même que l'utilisation d'une liste à gauche est rare et peu pythonique,
- les deux termes aient la même longueur - en tous cas avec les concepts que l'on a vus jusqu'ici, mais voir aussi plus bas l'utilisation de **arg* avec le *extended unpacking*.

La plupart du temps le terme de gauche est écrit comme un tuple. C'est pour cette raison que les deux termes *tuple unpacking* et *sequence unpacking* sont en vigueur.

47.1.4 La façon *pythonique* d'échanger deux variables

Une caractéristique intéressante de l'affectation par *sequence unpacking* est qu'elle est sûre; on n'a pas à se préoccuper d'un éventuel ordre d'évaluation, les valeurs **à droite** de l'affectation sont **toutes** évaluées en premier, et ainsi on peut par exemple échanger deux variables comme ceci :

```
a = 1
b = 2
a, b = b, a
print('a', a, 'b', b)
```

```
a 2 b 1
```

47.1.5 *Extended unpacking*

Le *extended unpacking* a été introduit en python3; commençons par en voir un exemple :

```
reference = [1, 2, 3, 4, 5]
a, *b, c = reference
print(f"a={a} b={b} c={c}")
```

```
a=1 b=[2, 3, 4] c=5
```

Comme vous le voyez, le mécanisme ici est une extension de *sequence unpacking*; python vous autorise à mentionner **une seule fois**, parmi les variables qui apparaissent à gauche de l'affectation, une variable **précédée de `***`**, ici `*b`.

Cette variable est interprétée comme une **liste de longueur quelconque** des éléments de `reference`. On aurait donc aussi bien pu écrire :

```
reference = range(20)
a, *b, c = reference
print(f"a={a} b={b} c={c}")
```

```
a=0 b=[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] c=19
```

Ce trait peut s'avérer pratique, lorsque par exemple on s'intéresse seulement aux premiers éléments d'une structure :

```
# si on sait que data contient prenom, nom, et un nombre inconnu d'autres_
↳ informations
data = [ 'Jean', 'Dupont', '061234567', '12', 'rue du chemin vert', '57000', 'METZ'
↳ ', ]
# on peut utiliser la variable _ qui véhicule l'idée qu'on ne s'y intéresse pas_
↳ vraiment
prenom, nom, *_ = data
print(f"prenom={prenom} nom={nom}")
```

```
prenom=Jean nom=Dupont
```

47.2 Complément - niveau intermédiaire

On a vu les principaux cas d'utilisation de la *sequence unpacking*, voyons à présent quelques subtilités.

47.2.1 Plusieurs occurrences d'une même variable

On peut utiliser **plusieurs fois** la même variable dans la partie gauche de l'affectation.

```
# ceci en toute rigueur est legal
# mais en pratique on évite de le faire
entree = [1, 2, 3]
a, a, a = entree
print(f"a = {a}")
```

```
a = 3
```

Attention toutefois, comme on le voit ici, python **n'impose pas** que les différentes occurrences de `a` correspondent à des valeurs identiques (en langage savant, on dirait que cela ne permet pas de faire de l'unification). De manière beaucoup plus pragmatique, l'interpréteur se contente de faire comme s'il faisait l'affectation plusieurs fois de gauche à droite, c'est-à-dire comme s'il faisait :

```
a = 1; a = 2; a = 3
```

Cette technique n'est utilisée en pratique que pour les parties de la structure dont on n'a que faire dans le contexte. Dans ces cas-là, il arrive qu'on utilise le nom de variable `_`, dont on rappelle qu'il est légal, ou tout autre nom comme `ignored` pour manifester le fait que cette partie de la structure ne sera pas utilisée, par exemple :

```
entree = [1, 2, 3]

_, milieu, _ = entree
print('milieu', milieu)

ignored, ignored, right = entree
print('right', right)
```

```
milieu 2
right 3
```

47.2.2 En profondeur

Le *sequence unpacking* ne se limite pas au premier niveau dans les structures, on peut extraire des données plus profondément imbriquées dans la structure de départ ; par exemple avec en entrée la liste :

```
structure = ['abc', [(1, 2), ([3], 4)], 5]
```

Si on souhaite extraire la valeur qui se trouve à l'emplacement du 3, on peut écrire :

```
(a, (b, ((trois), c)), d) = structure
print('trois', trois)
```

```
trois 3
```

Ou encore, sans doute un peu plus lisible :

```
(a, (b, ([trois], c)), d) = structure
print('trois', trois)
```

```
trois 3
```

Naturellement on aurait aussi bien pu écrire ici quelque chose comme :

```
trois = structure[1][1][0][0]
print('trois', trois)
```

```
trois 3
```

Affaire de goût évidemment. Mais n’oublions pas une des phrases du zen de python *Flat is better than nested*, ce qui veut dire que ce n’est pas parce que vous pouvez faire des structures imbriquées complexes que vous devez le faire. Bien souvent, cela rend la lecture et la maintenance du code complexe, j’espère que l’exemple précédent vous en a convaincu.

47.2.3 *Extended unpacking* et profondeur

On peut naturellement ajouter de l’*extended unpacking* à n’importe quel étage d’un *unpacking* imbriqué.

```
# un exemple très alambiqué avec plusieurs variables *extended
tree = [1, 2, [(3, 33, 'three', 'thirty-three')], ([4, 44, ('forty', 'forty-four
↪')])]
*_ , ((_, *x3, _),), (*_, x4) = tree
print(f"x3={x3}, x4={x4}")
```

```
x3=[33, 'three'], x4=('forty', 'forty-four')
```

Dans ce cas, la limitation d’avoir une seule variable de la forme **extended* s’applique toujours, naturellement, mais à chaque niveau dans l’imbrication, comme on le voit sur cet exemple.

47.3 Pour en savoir plus

— Le PEP (en anglais) qui introduit le **extended unpacking**

Plusieurs variables dans une boucle `for`

48.1 Complément - niveau basique

Nous avons vu précédemment (séquence “Les tuples”, complément “Sequence unpacking”) la possibilité d’affecter plusieurs variables à partir d’un seul objet, comme ceci :

```
item = (1, 2)
a, b = item
print(f"a={a} b={b}")
```

```
a=1 b=2
```

D’une façon analogue, il est possible de faire une boucle `for` qui itère sur **une seule** liste mais qui *agit* sur **plusieurs variables**, comme ceci :

```
entrees = [(1, 2), (3, 4), (5, 6)]
for a, b in entrees:
    print(f"a={a} b={b}")
```

```
a=1 b=2
a=3 b=4
a=5 b=6
```

À chaque itération, on trouve dans `entree` un tuple (d’abord `(1, 2)`, puis à l’iteration suivante `(3, 4)`, etc..); à ce stade les variables `a` et `b` vont être affectées à, respectivement, le premier et le deuxième élément du tuple, exactement comme dans le *sequence unpacking*. Cette mécanique est massivement utilisée en python.

48.2 Complément - niveau intermédiaire

48.2.1 La fonction `zip`

Voici un exemple très simple qui utilise la technique qu’on vient de voir.

Imaginons qu’on dispose de deux listes de longueurs égales, dont on sait que les entrées correspondent une à une, comme par exemple :

```
villes = ["Paris", "Nice", "Lyon"]
populations = [2*10**6, 4*10**5, 10**6]
```

Afin d'écrire facilement un code qui « associe » les deux listes entre elles, python fournit une fonction *built-in* baptisée `zip`; voyons ce qu'elle peut nous apporter sur cet exemple :

```
list(zip(villes, populations))
```

```
[('Paris', 2000000), ('Nice', 400000), ('Lyon', 1000000)]
```

On le voit, on obtient en retour une liste composée de tuples. On peut à présent écrire une boucle `for` comme ceci :

```
for ville, population in zip(villes, populations):
    print(population, "habitants à", ville)
```

```
2000000 habitants à Paris
400000 habitants à Nice
1000000 habitants à Lyon
```

Qui est, nous semble-t-il, beaucoup plus lisible que ce que l'on serait amené à écrire avec des langages plus traditionnels.

Tout ceci se généralise naturellement à plus de deux variables.

```
for i, j, k in zip(range(3), range(100, 103), range(200, 203)):
    print(f"i={i} j={j} k={k}")
```

```
i=0 j=100 k=200
i=1 j=101 k=201
i=2 j=102 k=202
```

Remarque : lorsqu'on passe à `zip` des listes de tailles différentes, le résultat est tronqué, c'est l'entrée **de plus petite taille** qui détermine la fin du parcours.

```
# on n'itère que deux fois
# car le premier argument de zip est de taille 2
for units, tens in zip([1, 2], [10, 20, 30, 40]):
    print(units, tens)
```

```
1 10
2 20
```

48.2.2 La fonction `enumerate`

Une autre fonction très utile permet d'itérer sur une liste avec l'indice dans la liste, il s'agit de `enumerate` :

```
for i, ville in enumerate(villes):
    print(i, ville)
```

```
0 Paris
1 Nice
2 Lyon
```

Cette forme est **plus simple** et **plus lisible** que les formes suivantes qui sont équivalentes, mais qui ne sont pas pythoniques :

```
for i in range(len(villes)):
    print(i, villes[i])
```

```
0 Paris
1 Nice
2 Lyon
```

```
for i, ville in zip(range(len(villes)), villes):
    print(i, ville)
```

```
0 Paris
1 Nice
2 Lyon
```


49.1 Exercice - niveau basique

49.1.1 Calcul du nombre de lignes, de mots et de caractères

```
# chargement de l'exercice
from corrections.exo_comptage import exo_comptage
```

On se propose d'écrire une * moulinette* qui annote un fichier avec des nombres de lignes, de mots et de caractères.

Le but de l'exercice est d'écrire une fonction `comptage` :

- qui prenne en argument un nom de fichier d'entrée (on suppose qu'il existe) et un nom de fichier de sortie (on suppose qu'on a le droit de l'écrire);
- le fichier d'entrée est supposé encodé en UTF-8;
- le fichier d'entrée est laissé intact;
- pour chaque ligne en entrée, le fichier de sortie comporte une ligne qui donne le numéro de ligne, le nombre de mots (**séparés par des espaces**), le nombre de caractères (y compris la fin de ligne), et la ligne d'origine.

```
# un exemple de ce qui est attendu
exo_comptage.example()
```

```
# votre code
def comptage(in_filename, out_filename):
    "votre code"
```

N'oubliez pas de vérifier que vous ajoutez bien les **fins de ligne**, car la vérification automatique est pointilleuse (elle utilise l'opérateur `==`), et rejettera votre code si vous ne produisez pas une sortie rigoureusement similaire à ce qui est attendu.

```
# pour vérifier votre code
# voyez aussi un peu plus bas, une cellule d'aide au debugging
```

```
exo_comptage.correction(comptage)
```

La méthode `debug` applique votre fonction au premier fichier d'entrée, et affiche le résultat comme dans l'exemple ci-dessus.

```
# debugging
exo_comptage.debug(comptage)
```

49.1.2 Accès aux fichiers d'exemples

Vous pouvez télécharger les fichiers d'exemples :

- [Romeo and Juliet](#)
- [Lorem Ipsum](#)
- « Une charogne » en utf-8

Pour les courageux, je vous donne également « Une charogne » en [Iso-latin-15](#), qui contient le même texte que « Une charogne », mais encodé en iso-latin-15.

Ce dernier fichier n'est pas à prendre en compte dans la version basique de l'exercice, mais vous pourrez vous rendre compte par vous même, au cas où cela ne serait pas clair encore pour vous, qu'il n'est pas facile d'écrire une fonction `comptage` qui devine l'encodage, c'est-à-dire qui fonctionne correctement avec des entrées indifféremment en unicode ou isolatin, sans que cet encodage soit passé en paramètre à `comptage`.

Sequence unpacking

50.1 Exercice - niveau basique

```
# chargeons l'exercice
from corrections.exo_surgery import exo_surgery
```

Cet exercice consiste à écrire une fonction `surgery`, qui prend en argument une liste, et qui retourne la **même** liste **modifiée** comme suit :

- si la liste est de taille 0 ou 1, elle n'est pas modifiée,
- si la liste est de taille paire, on intervertit les deux premiers éléments de la liste,
- si elle est de taille impaire, on intervertit les deux derniers éléments.

```
# voici quelques exemples de ce qui est attendu
exo_surgery.example()
```

```
# écrivez votre code
def surgery(liste):
    "<votre_code>"
```

```
# pour le vérifier, évaluez cette cellule
exo_surgery.correction(surgery)
```


51.1 Complément - niveau basique

Ce document résume les opérations courantes disponibles sur le type `dict`. On rappelle que le type `dict` est un type **mutable**.

51.1.1 Création en extension

On l'a vu, la méthode la plus directe pour créer un dictionnaire est en extension comme ceci :

```
annuaire = {'marc': 35, 'alice': 30, 'eric': 38}
print(annuaire)
```

```
{'marc': 35, 'alice': 30, 'eric': 38}
```

51.1.2 Création - la fonction dict

Comme pour les fonctions `int` ou `list`, la fonction `dict` est une fonction de construction de dictionnaire - on dit un constructeur. On a vu aussi dans la vidéo qu'on peut utiliser ce constructeur à base d'une liste de tuples (`clé, valeur`)

```
# le paramètre de la fonction dict est
# une liste de couples (clé, valeur)
annuaire = dict([('marc', 35), ('alice', 30), ('eric', 38)])
print(annuaire)
```

```
{'marc': 35, 'alice': 30, 'eric': 38}
```

Remarquons qu'on peut aussi utiliser cette autre forme d'appel à `dict` pour un résultat équivalent

```
annuaire = dict(marc=35, alice=30, eric=38)
print(annuaire)
```

```
{'marc': 35, 'alice': 30, 'eric': 38}
```

Remarquez ci-dessus l'absence de quotes autour des clés comme `marc`. Il s'agit d'un cas particulier de passage d'arguments que nous expliciterons plus longuement en fin de semaine 4.

51.1.3 Accès atomique

Pour accéder à la valeur associée à une clé, on utilise la notation à base de crochets `[]`

```
print('la valeur pour marc est', annuaire['marc'])
```

```
la valeur pour marc est 35
```

Cette forme d'accès ne fonctionne que si la clé est effectivement présente dans le dictionnaire. Dans le cas contraire, une exception `KeyError` est levée. Aussi si vous n'êtes pas sûr que la clé soit présente, vous pouvez utiliser la méthode `get` qui accepte une valeur par défaut :

```
print('valeur pour marc', annuaire.get('marc', 0))
print('valeur pour inconnu', annuaire.get('inconnu', 0))
```

```
valeur pour marc 35
valeur pour inconnu 0
```

Le dictionnaire est un type **mutable**, et donc on peut **modifier la valeur** associée à une clé :

```
annuaire['eric'] = 39
print(annuaire)
```

```
{'marc': 35, 'alice': 30, 'eric': 39}
```

Où encore, exactement de la même façon, **ajouter une entrée** :

```
annuaire['bob'] = 42
print(annuaire)
```

```
{'marc': 35, 'alice': 30, 'eric': 39, 'bob': 42}
```

Enfin pour **détruire une entrée**, on peut utiliser l'instruction `del` comme ceci :

```
# pour supprimer la clé 'marc' et donc sa valeur aussi
del annuaire['marc']
print(annuaire)
```

```
{'alice': 30, 'eric': 39, 'bob': 42}
```

Pour savoir si une clé est présente ou non, il est conseillé d'utiliser l'opérateur d'appartenance `in` comme ceci :

```
# forme recommandée
print('john' in annuaire)
```

```
False
```

51.1.4 Parcourir toutes les entrées

La méthode la plus fréquente pour parcourir tout un dictionnaire est à base de la méthode `items`; voici par exemple comment on pourrait afficher le contenu :

```
for nom, age in annuaire.items():
    print(f"{nom}, age {age}")
```

```
alice, age 30
eric, age 39
bob, age 42
```

On remarque d'abord que les entrées sont listées dans le désordre, plus précisément, il n'y a pas de notion d'ordre dans un dictionnaire ; ceci est dû à l'action de la fonction de hachage, que nous avons vue dans la vidéo précédente.

On peut obtenir séparément la liste des clés et des valeurs avec :

```
for clé in annuaire.keys():
    print(clé)
```

```
alice
eric
bob
```

```
for valeur in annuaire.values():
    print(valeur)
```

```
30
39
42
```

51.1.5 La fonction len

On peut comme d'habitude obtenir la taille d'un dictionnaire avec la fonction len :

```
print(f"{len(annuaire)} entrées dans annuaire")
```

```
3 entrées dans annuaire
```

51.1.6 Pour en savoir plus sur le type dict

Pour une liste exhaustive reportez-vous à la page de la documentation python ici

<https://docs.python.org/3/library/stdtypes.html#mapping-types-dict>

51.2 Complément - niveau intermédiaire

51.2.1 La méthode update

On peut également modifier un dictionnaire avec le contenu d'un autre dictionnaire avec la méthode update :

```
print(f"avant: {list(annuaire.items())}")
```

```
avant: [('alice', 30), ('eric', 39), ('bob', 42)]
```

```
annuaire.update({'jean':25, 'eric':70})
list(annuaire.items())
```

```
[('alice', 30), ('eric', 70), ('bob', 42), ('jean', 25)]
```

51.2.2 collections.OrderedDict : dictionnaire et ordre d'insertion

Attention : un dictionnaire est **non ordonné** ! Il ne se souvient pas de l'ordre dans lequel les éléments ont été insérés. C'était particulièrement visible dans les versions de python jusque 3.5 :

```
%%python2

# cette cellule utilise python-2.7 pour illustrer le fait
# que les dictionnaires ne sont pas ordonnes

d = {'c' : 3, 'b' : 1, 'a' : 2}
for k, v in d.items():
    print k, v
```

```
a 2
c 3
b 1
```

En réalité, et depuis la version 3.6 de python, il se trouve qu'**incidemment** l'implémentation CPython (la plus répandue donc) a été modifiée, et maintenant on peut avoir l'**impression** que les dictionnaires sont ordonnés :

```
d = {'c' : 3, 'b' : 1, 'a' : 2}
for k, v in d.items():
    print(k, v)
```

```
c 3
b 1
a 2
```

Il faut insister sur le fait qu'il s'agit d'un **détail d'implémentation**, et que vous ne devez pas écrire du code qui suppose que les dictionnaires sont ordonnés.

Si vous avez besoin de dictionnaires qui sont **garantis** ordonnés, voyez dans *le module* `'collections'` <<https://docs.python.org/3/library/collections.html>> la classe `OrderedDict` <<https://docs.python.org/3/library/collections.html#collections.OrderedDict>> qui est une customisation (une sous-classe) du type `dict`, qui cette fois possède cette bonne propriété :

```
from collections import OrderedDict
d = OrderedDict()
for i in ['a', 7, 3, 'x']:
    d[i] = i
for k, v in d.items():
    print('OrderedDict', k, v)
```

```
OrderedDict a a
OrderedDict 7 7
OrderedDict 3 3
OrderedDict x x
```

51.2.3 collections.defaultdict : initialisation automatique

Imaginons que vous devez gérer un dictionnaire dont les valeurs sont des listes, et que votre programme ajoute des valeurs au fur et à mesure dans ces listes.

Avec un dictionnaire de base, cela peut vous amener à écrire un code qui ressemble à ceci :

```
# on lit dans un fichier des couples (x, y)

tuples = [
    (1, 2),
```



```
(2, 1),
(1, 3),
(2, 4),
]
```

```
# et on veut construire un dictionnaire
# x -> [ liste de tous les y connectés à x]
resultat = {}

for x, y in tuples:
    if x not in resultat:
        resultat[x] = []
    resultat[x].append(y)

for key, value in resultat.items():
    print(key, value)
```

```
1 [2, 3]
2 [1, 4]
```

Cela fonctionne, mais n'est pas très élégant. Pour simplifier ce type de traitements, vous pouvez utiliser defaultdict, une sous-classe de dict dans le module collections :

```
from collections import defaultdict

# on indique que les valeurs doivent être créés à la volée
# en utilisant la fonction list
resultat = defaultdict(list)

# du coup plus besoin de vérifier la présence de la clé
for x, y in tuples:
    resultat[x].append(y)

for key, value in resultat.items():
    print(key, value)
```

```
1 [2, 3]
2 [1, 4]
```

Cela fonctionne aussi avec le type int, lorsque vous voulez par exemple compter des occurrences :

```
compteurs = defaultdict(int)

phrase = "une phrase dans laquelle on veut compter les caractères"

for c in phrase:
    compteurs[c] += 1

sorted(compteurs.items())
```

```
[(' ', 8),
('a', 5),
('c', 3),
('d', 1),
('e', 8),
('h', 1),
('l', 4),
('m', 1),
('n', 3),
('o', 2),
('p', 2),
```

```
('q', 1),
('r', 4),
('s', 4),
('t', 3),
('u', 3),
('v', 1),
('è', 1)]
```

Signalons enfin une fonctionnalité un peu analogue, quoiqu'un peut moins élégante à mon humble avis, mais qui est présente avec les dictionnaires `dict` standard. Il s'agit de la méthode `'setdefault'` <<https://docs.python.org/3/library/stdtypes.html#dict.setdefault>> qui permet, en un seul appel, de retourner la valeur associée à une clé et de créer cette clé au besoin, c'est-à-dire si elle n'est pas encore présente :

```
print('avant', annuaire)
# ceci sera sans effet car eric est déjà présent
print('set_default eric', annuaire.setdefault('eric', 50))
# par contre ceci va insérer une entrée dans le dictionnaire
print('set_default inconnu', annuaire.setdefault('inconnu', 50))
# comme on le voit
print('après', annuaire)
```

```
avant {'alice': 30, 'eric': 70, 'bob': 42, 'jean': 25}
set_default eric 70
set_default inconnu 50
après {'alice': 30, 'eric': 70, 'bob': 42, 'jean': 25, 'inconnu': 50}
```

Notez bien que `setdefault` peut éventuellement créer une entrée mais ne **modifie jamais** la valeur associée à une clé déjà présente dans le dictionnaire, comme le nom le suggère d'ailleurs.

51.3 Complément - niveau avancé

Pour bien appréhender les dictionnaires, il nous faut souligner certaines particularités, à propos de la valeur de retour des méthodes comme `items()`, `keys()` et `values()`.

Ce sont des objets itérables

Les méthodes `items()`, `keys()` et `values()` ne retournent pas des listes (comme c'était le cas avec python2), mais des **objets itérables** :

```
d = {'a' : 1, 'b' : 2}
keys = d.keys()
keys
```

```
dict_keys(['a', 'b'])
```

comme ce sont des itérables, on peut naturellement faire un `for` avec, on l'a vu

```
for key in keys:
    print(key)
```

```
a
b
```

et un test d'appartenance avec `in`

```
print('a' in keys)
```

```
True
```

```
print('x' in keys)
```

```
False
```

Mais ce ne sont pas des listes

```
isinstance(keys, list)
```

```
False
```

Ce qui signifie qu'on n'a **pas alloué de mémoire** pour stocker toutes les clés, mais seulement un objet qui ne prend pas de place, ni de temps à construire :

```
# construisons un dictionnaire
# pour anticiper un peu sur la compréhension de dictionnaire

big_dict = {k : k**2 for k in range(1_000_000)}
```

```
%%timeit -n 10000
# créer un objet vue est très rapide
big_keys = big_dict.keys()
```

```
119 ns ± 7.55 ns per loop (mean ± std. dev. of 7 runs, 10000 loops each)
```

```
# on répète ici car timeit travaille dans un espace qui lui est propre
# et donc on n'a pas défini big_keys pour notre interpréteur
big_keys = big_dict.keys()
```

```
%%timeit -n 20
# si on devait vraiment construire la liste ce serait beaucoup plus long
big_lkeys = list(big_keys)
```

```
20 ms ± 471 µs per loop (mean ± std. dev. of 7 runs, 20 loops each)
```

En fait ce sont des vues

Une autre propriété un peu inattendue de ces objets, c'est que **ce sont des vues** ; ce qu'on veut dire par là (pour ceux qui connaissent, cela fait référence à la notion de vue dans les bases de données) c'est que la vue *voit* les changements fait sur l'objet dictionnaire *même après sa création* :

```
d = {'a' : 1, 'b' : 2}
keys = d.keys()
```

```
# sans surprise, il y a deux clés dans keys
for k in keys:
    print(k)
```

```
a
b
```

```
# mais si maintenant j'ajoute un objet au dictionnaire
d['c'] = 3
# alors on va 'voir' cette nouvelle clé à partir de l'objet keys
# qui pourtant est inchangé
for k in keys:
    print(k)
```

```
a  
b  
c
```

Reportez vous à [la section sur les vues de dictionnaires](#) pour plus de détails.

python2

Ceci est naturellement en fort contraste avec tout ce qui se passait en python2, où l'on avait des méthodes distinctes, par exemple `keys()`, `iterkeys()` et `viewkeys()`, selon le type d'objets que l'on souhaitait construire.

52.1 Complément - niveau intermédiaire

Nous avons vu comment manipuler un dictionnaire, il nous reste à voir un peu plus en détail les contraintes qui sont mises par le langage sur ce qui peut servir de clé dans un dictionnaire. On parle dans ce complément spécifiquement des clefs contruites à partir des types `builtin`. Le cas de vos propres classes utilisées comme clefs de dictionnaires n'est pas abordé dans ce complément.

52.1.1 Une clé doit être immuable

Si vous vous souvenez de la vidéo sur les tables de hash, la mécanique interne du dictionnaire repose sur le calcul, à partir de chaque clé, d'une fonction de hachage.

C'est-à-dire que pour simplifier, on localise la présence d'une clé en calculant d'abord

$f(\text{clé}) = \text{hash}$

puis on poursuit la recherche en utilisant `hash` comme indice dans le tableau contenant les couples (clé, valeur).

On le rappelle, c'est cette astuce qui permet de réaliser les opérations sur les dictionnaires en temps constant - c'est-à-dire indépendamment du nombre d'éléments.

Cependant, pour que ce mécanisme fonctionne, il est indispensable que **la valeur de la clé reste inchangée** pendant la durée de vie du dictionnaire. Sinon, bien entendu, on pourrait avoir le scénario suivant :

- on range un tuple `(clef, valeur)` à un premier indice $f(\text{clef}) = \text{hash}_1$
- on modifie la valeur de `clef` qui devient `clef'`
- on recherche notre valeur à l'indice $f(\text{clef}') = \text{hash}_2 \neq \text{hash}_1$

et donc avec ces hypothèses on n'a plus la garantie de bon fonctionnement de la logique.

52.1.2 Une clé doit être globalement immuable

Nous avons depuis le début du cours longuement insisté sur le caractère mutable ou immuable des différents types prédéfinis de python. Vous devez donc à présent avoir au moins en partie ce tableau en tête :

| Type | Mutable ? |
|---------------|-----------|
| int, float | immuable |
| complex, bool | immuable |
| str | immuable |
| list | mutable |
| dict | mutable |
| set | mutable |
| frozenset | immuable |

Le point important ici, est qu'il **ne suffit pas**, pour une clé, d'être **de type immuable**.

On peut le voir sur un exemple très simple ; donnons nous donc un dictionnaire

```
d = {}
```

Et commençons avec un objet de type immuable, un tuple d'entiers

```
bonne_cle = (1, 2)
```

Cet objet est non seulement **de type immuable**, mais tous ses composants et sous-composants sont **immuables**, on peut donc l'utiliser comme clé dans le dictionnaire

```
d[bonne_cle] = "pas de probleme ici"
print(d)
```

```
{(1, 2): 'pas de probleme ici'}
```

Si à présent on essaie d'utiliser comme clé un tuple qui contient une liste :

```
mauvaise_cle = (1, [1, 2])
```

Il se trouve que cette clé, **bien que de type immuable**, peut être **indirectement modifiée** puisque :

```
mauvaise_cle[1].append(3)
print(mauvaise_cle)
```

```
(1, [1, 2, 3])
```

Et c'est pourquoi on ne peut pas utiliser cet objet comme clé dans le dictionnaire

```
# provoque une exception
d[mauvaise_cle] = 'on ne peut pas faire ceci'
```

```
-----
TypeError                                Traceback (most recent call last)
<ipython-input-6-08a3625a4eed> in <module>()
      1 # provoque une exception
----> 2 d[mauvaise_cle] = 'on ne peut pas faire ceci'

TypeError: unhashable type: 'list'
```

Pour conclure, il faut retenir qu'un objet n'est éligible pour être utilisé comme clé que s'il est **composé de types immuables du haut en bas** de la structure de données.

La raison d'être principale du type `tuple`, que nous avons vu la semaine passée, et du type `frozenset`, que nous verrons très prochainement, est précisément de construire de tels objets globalement immuables.

52.1.3 Épilogue

Tout ceci est valable pour les types *builtin*. Nous verrons que pour les types définis par l'utilisateur - les classes donc - que nous effleurons à la fin de cette semaine et que nous étudions plus en profondeur en semaine 6, c'est un autre mécanisme qui est utilisé pour calculer la clé de hachage d'une instance de classe.

Gérer des enregistrements

53.1 Complément - niveau intermédiaire

53.1.1 Implémenter un enregistrement comme un dictionnaire

Il nous faut faire le lien entre dictionnaire python et la notion d'enregistrement, c'est-à-dire une donnée composite qui contient plusieurs champs. (À cette notion correspond, selon les langages, ce qu'on appelle un `struct` ou un `record`)

Imaginons qu'on veuille manipuler un ensemble de données concernant des personnes ; chaque personne est supposée avoir un nom, un age et une adresse mail.

Il est possible, et assez fréquent, d'utiliser le dictionnaire comme support pour modéliser ces données comme ceci :

```
personnes = [
    {'nom': 'pierre', 'age': 25, 'email': 'pierre@foo.com'},
    {'nom': 'paul', 'age': 18, 'email': 'paul@bar.com'},
    {'nom': 'jacques', 'age': 52, 'email': 'jacques@cool.com'},
]
```

Bon, très bien, nous avons nos données, il est facile de les utiliser.

Par exemple, pour l'anniversaire de pierre on fera :

```
personnes[0]['age'] += 1
```

Ce qui nous donne

```
for personne in personnes:
    print(10*"=")
    for info, valeur in list(personne.items()):
        print(f"{info} -> {valeur}")
```

```
=====
nom -> pierre
age -> 26
email -> pierre@foo.com
=====
```

```

nom -> paul
age -> 18
email -> paul@bar.com
=====
nom -> jacques
age -> 52
email -> jacques@cool.com

```

53.1.2 Un dictionnaire pour indexer les enregistrements

Cela dit, il est bien clair que cette façon de faire n'est pas très pratique ; pour marquer l'anniversaire de pierre on ne sait bien entendu pas que son enregistrement est le premier dans la liste. C'est pourquoi il est plus adapté, pour modéliser ces informations, d'utiliser non pas une liste, mais à nouveau... un dictionnaire.

Si on imagine qu'on a commencé par lire ces données séquentiellement dans un fichier, et qu'on a calculé l'objet personnes comme la liste qu'on a vue ci-dessus, alors il est possible de construire un index de ces dictionnaires, (un dictionnaire de dictionnaires, donc).

C'est-à-dire, en anticipant un peu sur la construction de dictionnaires par compréhension :

```

# on crée un index permettant de retrouver rapidement
# une personne dans la liste
index_par_nom = {personne['nom']: personne for personne in personnes}

print("enregistrement pour pierre", index_par_nom['pierre'])

```

```

enregistrement pour pierre {'nom': 'pierre', 'age': 26, 'email': 'pierre@foo.com'}

```

Attardons nous un tout petit peu ; nous avons construit un dictionnaire par compréhension, en créant autant d'entrées que de personnes. Nous aborderons en détail la notion de compréhension de sets et de dictionnaires en semaine 5, donc si cette notation vous paraît étrange pour le moment, pas d'inquiétude.

Le résultat est donc un dictionnaire qu'on peut afficher comme ceci :

```

for nom, record in index_par_nom.items():
    print(f"Nom : {nom} -> enregistrement : {record}")

```

```

Nom : pierre -> enregistrement : {'nom': 'pierre', 'age': 26, 'email': 'pierre@foo.
↪com'}
Nom : paul -> enregistrement : {'nom': 'paul', 'age': 18, 'email': 'paul@bar.com'}
Nom : jacques -> enregistrement : {'nom': 'jacques', 'age': 52, 'email':
↪'jacques@cool.com'}

```

Dans cet exemple, le premier niveau de dictionnaire permet de trouver rapidement un objet à partir d'un nom ; dans le second niveau au contraire on utilise le dictionnaire pour implémenter un enregistrement, à la façon d'un struct en C.

53.1.3 Techniques similaires

Notons enfin qu'il existe aussi, en python, un autre mécanisme qui peut être utilisé pour gérer ce genre d'objets composites, ce sont les classes que nous verrons en semaine 6, et qui permettent de définir de nouveaux types plutôt que, comme nous l'avons fait ici, d'utiliser un type prédéfini. Dans ce sens, l'utilisation d'une classe permet davantage de souplesse, au prix de davantage d'effort.

53.2 Complément - niveau avancé

53.2.1 La même idée, mais avec une classe `Personne`

Je vais donner ici une implémentation du code ci-dessus, qui utilise une classe pour modéliser les personnes. Naturellement je n'entre pas dans les détails, que l'on verra en semaine 6, mais j'espère vous donner un aperçu des classes dans un usage réaliste, et vous montrer les avantages de cette approche.

Pour commencer je définis la classe `Personne`, qui va me servir à modéliser chaque personne.

```
class Personne:

    # le constructeur - vous ignorez le paramètre self,
    # on pourra construire une personne à partir de
    # 3 paramètres
    def __init__(self, nom, age, email):
        self.nom = nom
        self.age = age
        self.email = email

    # je définis cette méthode pour avoir
    # quelque chose de lisible quand je print()
    def __repr__(self):
        return f"{self.nom} ({self.age} ans) sur {self.email}"
```

Pour construire ma liste de personnes, je fais alors :

```
personnes2 = [
    Personne('pierre', 25, 'pierre@foo.com'),
    Personne('paul', 18, 'paul@bar.com'),
    Personne('jacques', 52, 'jacques@cool.com'),
]
```

Si je regarde un élément de la liste j'obtiens :

```
personnes2[0]
```

```
pierre (25 ans) sur pierre@foo.com
```

Je peux indexer tout ceci comme tout à l'heure, si j'ai besoin d'un accès rapide :

```
# je dois utiliser cette fois personne.nom et non plus personne['nom']
index2 = {personne.nom : personne for personne in personnes2}
```

Le principe ici est exactement identique à ce qu'on a fait avec le dictionnaire de dictionnaires, mais on a construit un dictionnaire d'instances.

Et de cette façon :

```
print(index2['pierre'])
```

```
pierre (25 ans) sur pierre@foo.com
```


54.1 Exercice - niveau basique

```
from corrections.exo_graph_dict import exo_graph_dict
```

On veut implémenter un petit modèle de graphes. Comme on a les données dans des fichiers, on veut analyser des fichiers d'entrée qui ressemblent à ceci :

```
!cat data/graph1.txt
```

```
s1 10 s2
s2 12 s3
s3 25 s1
s1 14 s3
```

qui signifierait :

- un graphe à 3 sommets $s1$, $s2$ et $s3$
- et 4 arêtes, par exemple une entre $s1$ et $s2$ de longueur 10.

On vous demande d'écrire une fonction qui lit un tel fichier texte, et construit (et retourne) un dictionnaire python qui représente ce graphe.

Dans cet exercice on choisit :

- de modéliser le graphe comme un dictionnaire indexé sur les (noms de) sommets,
- et chaque valeur est une liste de tuples de la forme (*suivant*, *longueur*), dans l'ordre d'apparition dans le fichier d'entrée.

```
# voici ce qu'on obtiendrait par exemple avec les données ci-dessus
exo_graph_dict.example()
```

```
# à vous de jouer
def graph_dict(filename):
    "votre code"
```

```
exo_graph_dict.correction(graph_dict)
```

Fusionner des données

55.1 Exercices

Cet exercice vient en deux versions, une de niveau basique et une de niveau intermédiaire.

La version basique est une application de la technique d'indexation qu'on a vue dans le complément « Gérer des enregistrements ». On peut très bien faire les deux versions dans l'ordre, une fois qu'on a fait la version basique on est en principe un peu plus avancé pour aborder la version intermédiaire.

55.1.1 Contexte

Nous allons commencer à utiliser des données un peu plus réalistes. Il s'agit de données obtenues auprès de [MarineTraffic](#) - et légèrement simplifiées pour les besoins de l'exercice. Ce site expose les coordonnées géographiques de bateaux observées en mer au travers d'un réseau de collecte de type *crowdsourcing*.

De manière à optimiser le volume de données à transférer, l'API de MarineTraffic offre deux modes pour obtenir les données

- **mode étendu** : chaque mesure (bateau x position x temps) est accompagnée de tous les détails du bateau (*id*, nom, pays de rattachement, etc.)
- **mode abrégé** : chaque mesure est uniquement attachée à l'*id* du bateau.

En effet, chaque bateau possède un identifiant unique qui est un entier, que l'on note *id*.

55.1.2 Chargement des données

Commençons par charger les données de l'exercice

```
from corrections.exo_marine_dict import extended, abbreviated
```

55.1.3 Format des données

Le format de ces données est relativement simple, il s'agit dans les deux cas d'une liste d'entrées - une par bateau.

Chaque entrée à son tour est une liste qui contient :

```
mode étendu: [id, latitude, longitude, date_heure, nom_bateau, code_pays, ...]
mode abrégé: [id, latitude, longitude, date_heure]
```

sachant que les entrées après le code pays dans le format étendu ne nous intéressent pas pour cet exercice.

```
# une entrée étendue est une liste qui ressemble à ceci
sample_extended_entry = extended[3]
print(sample_extended_entry)
```

```
[255801560, 49.3815, -4.412167, '2013-10-08T21:51:00', 'AUTOPRIDE', 'PT', '',
↪ 'ZEEBRUGGE']
```

```
# une entrée abrégée ressemble à ceci
sample_abbreviated_entry = abbreviated[0]
print(sample_abbreviated_entry)
```

```
[227254910, 49.91799, -5.315172, '2013-10-08T22:59:00']
```

On précise également que les deux listes `extended` et `abbreviated`

- possèdent exactement **le même nombre** d'entrées
- et correspondent **aux mêmes bateaux**
- mais naturellement **à des moments différents**
- et **pas forcément dans le même ordre**.

55.1.4 Exercice - niveau basique

```
# chargement de l'exercice
from corrections.exo_marine_dict import exo_index
```

But de l'exercice

On vous demande d'écrire une fonction `index` qui calcule, à partir de la liste des données étendues, un dictionnaire qui est :

- indexé par l'id de chaque bateau,
- et qui a pour valeur la liste qui décrit le bateau correspondant.

De manière plus imagée, si :

```
extended = [ bateau1, bateau2, ... ]
```

et si

```
bateau1 = [ id1, latitude, ... ]
```

on doit obtenir comme résultat de `index` un dictionnaire

```
{ id1 -> [ id_bateau1, latitude, ... ],
  id2 ...
}
```


Bref, on veut pouvoir retrouver les différents éléments de la liste `extended` par accès direct, en ne faisant qu'un seul *lookup* dans l'index.

```
# le résultat attendu
result_index = exo_index.resultat(extended)

# on en profite pour illustrer le module pprint
from pprint import pprint

# à quoi ressemble le résultat pour un bateau au hasard
for key, value in result_index.items():
    print("==== clé")
    pprint(key)
    print("==== valeur")
    pprint(value)
    break
```

```
==== clé
992271012
==== valeur
[992271012, 47.64744, -3.509282, '2013-10-08T21:50:00', 'PENMEN', 'FR', '', '']
```

Remarquez ci-dessus l'utilisation d'un utilitaire parfois pratique : le module `"pprint"` pour pretty-printer <<https://docs.python.org/3/library/pprint.html>>‘__.

Votre code

```
def index(extended):
    "<votre_code>"
```

Validation

```
exo_index.correction(index, abbreviated)
```

Vous remarquerez d'ailleurs que la seule chose qu'on utilise dans cet exercice, c'est que l'id des bateaux arrive en première position (dans la liste qui matérialise le bateau), aussi votre code doit marcher à l'identique avec les bateaux étendus :

```
exo_index.correction(index, extended)
```

55.1.5 Exercice - niveau intermédiaire

```
# chargement de l'exercice
from corrections.exo_marine_dict import exo_merge
```

But de l'exercice

On vous demande d'écrire une fonction `merge` qui fasse une consolidation des données, de façon à obtenir en sortie un dictionnaire :

```
id -> [ nom_bateau, code_pays, position_etendu, position_abrege ]
```

dans lequel les deux objets `position` sont tous les deux des tuples de la forme

```
(latitude, longitude, date_heure)
```

Voici par exemple un couple clé-valeur dans le résultat attendu.

```
# le résultat attendu
result_merge = exo_merge.resultat(extended, abbreviated)

# a quoi ressemble le résultat pour un bateau au hasard
from pprint import pprint
for key_value in result_merge.items():
    pprint(key_value)
    break
```

```
(992271012,
 [ 'PENMEN',
   'FR',
   (47.64744, -3.509282, '2013-10-08T21:50:00'),
   (47.64748, -3.509307, '2013-10-08T22:56:00')])
```

Votre code

```
def merge(extended, abbreviated):
    "votre code"
```

Validation

```
exo_merge.correction(merge, extended, abbreviated)
```

55.1.6 Les fichiers de données complets

Signalons enfin pour ceux qui sont intéressés que les données chargées dans cet exercice sont disponibles au format JSON - qui est précisément celui exposé par marinetransit.

Nous avons beaucoup simplifié les données d'entrée pour vous permettre une mise au point plus facile. Si vous voulez vous amuser à charger des données un peu plus significatives, sachez que

- vous avez accès aux fichiers de données plus complets :
 - `data/marine-e1-ext.json`
 - `data/marine-e1-abb.json`
- pour charger ces fichiers, qui sont donc au **format JSON**, la connaissance intime de ce format n'est pas nécessaire, on peut tout simplement utiliser le *module* `'json'` <<https://docs.python.org/3/library/json.html>>'. Voici le code utilisé dans l'exercice pour charger ces JSON en mémoire ; il utilise des notions que nous verrons dans les semaines à venir :

```
# load data from files
import json

with open("data/marine-e1-ext.json", encoding="utf-8") as feed:
    extended_full = json.load(feed)

with open("data/marine-e1-abb.json", encoding="utf-8") as feed:
    abbreviated_full = json.load(feed)
```

Une fois que vous avez un code qui fonctionne vous pouvez le lancer sur ces données plus copieuses en faisant

```
exo_merge.correction(merge, extended_full, abbreviated_full)
```


56.1 Complément - niveau basique

Ce document résume les opérations courantes disponibles sur le type `set`. On rappelle que le type `set` est un type **mutable**.

56.1.1 Création en extension

On crée un ensemble avec les accolades, comme les dictionnaires, mais sans utiliser le caractère `:`, et cela donne par exemple :

```
heteroclite = {'marc', 12, 'pierre', (1, 2, 3), 'pierre'}  
print(heteroclite)
```

```
{'marc', 12, (1, 2, 3), 'pierre'}
```

56.1.2 Création - la fonction `set`

Il devrait être clair à ce stade que, le nom du type étant `set`, la fonction `set` est un constructeur d'ensembles. On aurait donc aussi bien pu faire :

```
heteroclite2 = set(['marc', 12, 'pierre', (1, 2, 3), 'pierre'])  
print(heteroclite2)
```

```
{'marc', 12, (1, 2, 3), 'pierre'}
```

56.1.3 Créer un ensemble vide

Il faut remarquer que l'on ne peut pas créer un ensemble vide en extension. En effet :

```
type({})
```

```
dict
```

Ceci est lié à des raisons historiques, les ensembles n'ayant fait leur apparition que tardivement dans le langage en tant que citoyen de première classe.

Pour créer un ensemble vide, la pratique la plus courante est celle-ci :

```
ensemble_vide = set()
print(type(ensemble_vide))
```

```
<class 'set'>
```

Ou également, moins élégant mais que l'on trouve parfois dans du vieux code :

```
autre_ensemble_vide = set([])
print(type(autre_ensemble_vide))
```

```
<class 'set'>
```

56.1.4 Un élément dans un ensemble doit être globalement immuable

On a vu précédemment que les clés dans un dictionnaire doivent être globalement immuables. Pour exactement les mêmes raisons, les éléments d'un ensemble doivent aussi être globalement immuables.

```
# on ne peut pas insérer un tuple qui contient une liste
>>> ensemble = {(1, 2, [3, 4])}
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unhashable type: 'list'
```

Le type `set` étant lui-même mutable, on ne peut pas créer un ensemble d'ensembles :

```
>>> ensemble = {{1, 2}}
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unhashable type: 'set'
```

Et c'est une des raisons d'être du type `frozenset`.

56.1.5 Création - la fonction `frozenset`

Un `frozenset` est un ensemble qu'on ne peut pas modifier, et qui donc peut servir de clé dans un dictionnaire, ou être inclus dans un autre ensemble (mutable ou pas).

Il n'existe pas de raccourci syntaxique comme les `{ }` pour créer un ensemble immuable, qui doit être créé avec la fonction `frozenset`. Toutes les opérations documentées dans ce notebook, et qui n'ont pas besoin de modifier l'ensemble, sont disponibles sur un `frozenset`.

Parmi les fonctions exclues sur un `frozenset`, on peut citer : `update`, `pop`, `clear`, `remove` ou `discard`.

56.1.6 Opérations simples

```
# pour rappel
heteroclite
```

```
{(1, 2, 3), 12, 'marc', 'pierre'}
```

Test d'appartenance

```
(1, 2, 3) in heteroclite
```

```
True
```

Cardinal

```
len(heteroclite)
```

```
4
```

Manipulations

```
ensemble = {1, 2, 1}  
ensemble
```

```
{1, 2}
```

```
# pour nettoyer  
ensemble.clear()  
ensemble
```

```
set()
```

```
# ajouter un element  
ensemble.add(1)  
ensemble
```

```
{1}
```

```
# ajouter tous les elements d'un autre *ensemble*  
ensemble.update({2, (1, 2, 3), (1, 3, 5)})  
ensemble
```

```
{(1, 2, 3), (1, 3, 5), 1, 2}
```

```
# enlever un element avec discard  
ensemble.discard((1, 3, 5))  
ensemble
```

```
{(1, 2, 3), 1, 2}
```

```
# discard fonctionne même si l'élément n'est pas présent  
ensemble.discard('foo')  
ensemble
```

```
{(1, 2, 3), 1, 2}
```

```
# enlever un élément avec remove  
ensemble.remove((1, 2, 3))  
ensemble
```

```
{1, 2}
```

```
# contrairement à discard, l'élément doit être présent,
# sinon il y a une exception
try:
    ensemble.remove('foo')
except KeyError as e:
    print("remove a levé l'exception", e)
```

```
remove a levé l'exception 'foo'
```

La capture d'exception avec `try` et `except` sert à capturer une erreur d'exécution du programme (qu'on appelle exception) pour continuer le programme. Le but de cet exemple est simplement de montrer (d'une manière plus élégante que de voir simplement le programme planter avec une exception non capturée) que l'expression `ensemble.remove('foo')` génère une exception. Si ce concept vous paraît obscur, pas d'inquiétude, nous l'aborderons cette semaine et nous y reviendrons en détail en semaine 6.

```
# pop() ressemble à la méthode éponyme sur les listes
# sauf qu'il n'y a pas d'ordre dans un ensemble
while ensemble:
    element = ensemble.pop()
    print("element", element)
print("et bien sûr maintenant l'ensemble est vide", ensemble)
```

```
element 1
element 2
et bien sûr maintenant l'ensemble est vide set()
```

56.1.7 Opérations classiques sur les ensembles

Donnons-nous deux ensembles simples :

```
A2 = set([0, 2, 4, 6])
print('A2', A2)
A3 = set([0, 6, 3])
print('A3', A3)
```

```
A2 {0, 2, 4, 6}
A3 {0, 3, 6}
```

N'oubliez pas que les ensembles, comme les dictionnaires, ne sont **pas ordonnés**.

Remarques

- Les notations des opérateurs sur les ensembles rappellent les opérateurs « bit-à-bit » sur les entiers.
- Ces opérateurs sont également disponibles sous la forme de méthodes

Union

```
A2 | A3
```

```
{0, 2, 3, 4, 6}
```


Intersection

```
A2 & A3
```

```
{0, 6}
```

Différence

```
A2 - A3
```

```
{2, 4}
```

```
A3 - A2
```

```
{3}
```

Différence symétrique

On rappelle que $A \Delta B = (A - B) \cup (B - A)$

```
A2 ^ A3
```

```
{2, 3, 4}
```

56.1.8 Comparaisons

Ici encore on se donne deux ensembles :

```
superset = {0, 1, 2, 3}
print('superset', superset)
subset = {1, 3}
print('subset', subset)
```

```
superset {0, 1, 2, 3}
subset {1, 3}
```

Égalité

```
heteroclite == heteroclite2
```

```
True
```

Inclusion

```
subset <= superset
```

```
True
```

```
subset < superset
```

```
True
```

```
heteroclite < heteroclite2
```

```
False
```

Ensembles disjoints

```
heteroclite.isdisjoint(A3)
```

```
True
```

56.1.9 Pour en savoir plus

Reportez vous à [la section sur les ensembles](#) dans la documentation python.

57.1 Exercice - niveau basique

```
# charger l'exercice
from corrections.exo_read_set import exo_read_set
```

On se propose d'écrire une fonction `read_set` qui construit un ensemble à partir du contenu d'un fichier. Voici par exemple un fichier d'entrée

```
!cat data/setrefl.txt
```

```
4615
 12
9228
6158
12
```

`read_set` va prendre en argument un nom de fichier (vous pouvez supposer qu'il existe), enlever les espaces éventuels au début et à la fin de chaque ligne, et construire un ensemble de toutes les lignes ; par exemple :

```
exo_read_set.example()
```

| Appel | Résultat Attendu |
|---|---|
| <code>read_set("data/setrefl.txt")</code> | <code>{ "12", "4615", "6158", "9228" }</code> |

```
# écrivez votre code ici
def read_set(filename):
    "votre code"
```

```
# vérifiez votre code ici
exo_read_set.correction(read_set)
```

57.2 Deuxième partie - niveau basique

```
# la définition de l'exercice
from corrections.exo_read_set import exo_search_in_set
```

Ceci étant acquis, on veut écrire une deuxième fonction `search_in_set` qui prend en argument deux fichiers :

- `filename_reference` est le nom d'un fichier contenant des mots de reference,
- `filename` est le nom d'un fichier contenant des mots, dont on veut savoir s'ils sont ou non dans les références.

Pour cela `search_in_set` doit retourner une liste, contenant pour chaque ligne du fichier `reference` un tuple avec

- la ligne (sans les espaces de début et de fin, ni la fin de ligne)
- un booléen qui indique si ce mot est présent dans les références ou pas.

Par exemple :

```
!cat data/setrefl.txt
```

```
4615
12
9228
6158
12
```

```
!cat data/setsample1.txt
```

```
2048
8192
9228
2049
3
4
2053
2054
6158
4099
8
12
```

```
exo_search_in_set.example()
```

| Appel | Résultat Attendu |
|---|---|
| <code>search_in_set("data/setrefl.txt", "data/setsample1.txt")</code> | [("2048", False), ("8192", False), ("9228", True), ("2049", False), ("3", False), ("4", False), ("2053", False), ("2054", False), ("6158", True), ("4099", False), ("8", False), ("12", True)] |

```
# à vous
def search_in_set(filename_reference, filename):
    "votre code"
```

```
# vérifiez  
exo_search_in_set.correction(search_in_set)
```

Exercice sur les ensembles

58.1 Exercice - niveau intermédiaire

```
# chargement de l'exercice
from corrections.exo_marine_set import exo_diff
```

58.1.1 Les données

Nous reprenons le même genre de données marines en provenance de MarineTraffic que nous avons vues dans l'exercice précédent.

```
from corrections.exo_marine_set import abbreviated, extended
```

58.1.2 Rappels sur les formats

```
étendu: [id, latitude, longitude, date_heure, nom_bateau, code_pays...]
abrégé: [id, latitude, longitude, date_heure]
```

```
print(extended[0])
```

```
[304010873, 49.54708, -3.548188, '2013-10-08T21:51:00', 'A.B.LIVERPOOL', 'AG', '',
↪ 'SINES']
```

```
print(abbreviated[0])
```

```
[477346300, 49.00058, -5.503775, '2013-10-08T22:56:00']
```

58.1.3 But de l'exercice

```
# chargement de l'exercice
from corrections.exo_marine_set import exo_diff
```

Notez bien une différence importante avec l'exercice précédent : cette fois **il n'y a plus correspondance** entre les bateaux rapportés dans les données étendues et abrégées.

Le but de l'exercice est précisément d'étudier la différence, et pour cela on vous demande d'écrire une fonction

```
diff(extended, abbreviated)
```

qui retourne un tuple à trois éléments

- l'ensemble (set) des **noms** des bateaux présents dans `extended` mais pas dans `abbreviated`
- l'ensemble des **noms** des bateaux présents dans `extended` et dans `abbreviated`
- l'ensemble des **id** des bateaux présents dans `abbreviated` mais pas dans `extended` (par construction, les données ne nous permettent pas d'obtenir les noms de ces bateaux)

```
# le résultat attendu
result = exo_diff.resultat(extended, abbreviated)

# combien de bateaux sont concernés
def show_result(extended, abbreviated, result):
    """
    Affiche divers décomptes sur les arguments
    en entrée et en sortie de diff
    """
    print(10*'- ', "Les entrées")
    print(f"Dans extended: {len(extended)} entrées")
    print(f"Dans abbreviated: {len(abbreviated)} entrées")
    print(10*'- ', "Le résultat du diff")
    extended_only, both, abbreviated_only = result
    print(f"Dans extended mais pas dans abbreviated {len(extended_only)}")
    print(f"Dans les deux {len(both)}")
    print(f"Dans abbreviated mais pas dans extended {len(abbreviated_only)}")

show_result(extended, abbreviated, result)
```

```
----- Les entrées
Dans extended: 4 entrées
Dans abbreviated: 4 entrées
----- Le résultat du diff
Dans extended mais pas dans abbreviated 2
Dans les deux 2
Dans abbreviated mais pas dans extended 2
```

58.1.4 Votre code

```
def diff(extended, abbreviated):
    "<votre_code>"
```

58.1.5 Validation

```
exo_diff.correction(diff, extended, abbreviated)
```

58.1.6 Des fichiers de données plus réalistes

Comme pour l'exercice précédent, les données fournies ici sont très simplistes ; vous pouvez si vous le voulez essayer votre code avec des données (un peu) plus réalistes en chargeant des fichiers de données plus complets :

- data/marine-e2-ext.json
- data/marine-e2-abb.json

Ce qui donnerait en python :

```
# load data from files
import json

with open("data/marine-e2-ext.json", encoding="utf-8") as feed:
    extended_full = json.load(feed)

with open("data/marine-e2-abb.json", encoding="utf-8") as feed:
    abbreviated_full = json.load(feed)
```

```
# le résultat de votre fonction sur des données plus vastes
# attention que show_result fait des hypothèses sur le type de votre résultat
# aussi si vous essayez d'exécuter ceci avec comme fonction diff
# la version vide qui est dans le notebook original
# cela peut provoquer une exception
diff_full = diff(extended_full, abbreviated_full)
show_result(extended_full, abbreviated_full, diff_full)
```

Je signale enfin à propos de ces données plus complètes que :

- on a supprimé les entrées correspondants à des bateaux différents mais de même nom ; cette situation peut arriver dans la réalité (c'est pourquoi d'ailleurs les bateaux ont un *id*) mais ici ce n'est pas le cas.
- il se peut par contre qu'un même bateau fasse l'objet de plusieurs mesures dans `extended` et/ou dans `abbreviated`.

```
try .. else .. finally
```

59.1 Complément - niveau intermédiaire

L'instruction `try` est généralement assortie d'une ou plusieurs clauses `except`, comme on l'a vu dans la vidéo.

Sachez qu'on peut aussi utiliser - après toutes les clauses `except` - une clause

- `else`, qui va être exécutée si aucune exception n'est attrapée, et/ou une clause
- `finally` qui sera alors exécutée quoi qu'il arrive.

Voyons cela sur des exemples.

59.1.1 `finally`

C'est sans doute `finally` qui est la plus utile de ces deux clauses, car elle permet de faire un nettoyage **dans tous les cas de figure** - de ce point de vue, cela rappelle un peu les *context managers*.

Et par exemple, comme avec les *context managers*, une fonction peut faire des choses même après un `return`.

```
# une fonction qui fait des choses après un return
def return_with_finally(number):
    try:
        return 1/number
    except ZeroDivisionError as e:
        print(f"OOPS, {type(e)}, {e}")
        return "zero-divide"
    finally:
        print("on passe ici même si on a vu un return")
```

```
# sans exception
return_with_finally(1)
```

```
on passe ici même si on a vu un return
```

```
1.0
```

```
# avec exception
return_with_finally(0)
```

```
OOPS, <class 'ZeroDivisionError'>, division by zero
on passe ici même si on a vu un return
```

```
'zero-divide'
```

59.1.2 else

La logique ici est assez similaire, sauf que le code du `else` n'est exécutée que dans le cas où aucune exception n'est attrapée.

En première approximation, on pourrait penser que c'est équivalent de mettre du code dans la clause `else` ou à la fin de la clause `try`. En fait il y a une différence subtile :

The use of the “else” clause is better than adding additional code to the “try” clause because it avoids accidentally catching an exception that wasn't raised by the code being protected by the “try”... “except” statement.

Dit autrement, si le code dans la clause `else` lève une exception, celle-ci ne sera **pas attrapée** par le `try` courant, et sera donc propagée.

Voici un exemple rapidement, en pratique on rencontre assez peu souvent une clause `else` dans un `try`.

```
# pour montrer la clause else dans un usage banal
def function_with_else(number):
    try:
        x = 1/number
    except ZeroDivisionError as e:
        print(f"OOPS, {type(e)}, {e}")
    else:
        print("on passe ici seulement avec un nombre non nul")
    return 'something else'
```

```
# sans exception
function_with_else(1)
```

```
on passe ici seulement avec un nombre non nul
```

```
'something else'
```

```
# avec exception
function_with_else(0)
```

```
OOPS, <class 'ZeroDivisionError'>, division by zero
```

```
'something else'
```

Remarquez que `else` ne présente pas cette particularité de « traverser » le `return`, qu'on a vu avec `finally` :

```
# la clause else ne traverse pas les return
def return_with_else(number):
    try:
        return 1/number
    except ZeroDivisionError as e:
```

```
print(f"OOPS, {type(e)}, {e}")
return("zero-divide")
else:
    print("on ne passe jamais ici à cause des return")
```

```
# sans exception
return_with_else(1)
```

```
1.0
```

```
# avec exception
return_with_else(0)
```

```
OOPS, <class 'ZeroDivisionError'>, division by zero
```

```
'zero-divide'
```

59.1.3 Pour en savoir plus

Voyez le [tutorial sur les exceptions](#) dans la documentation officielle.

60.1 Complément - niveau basique

```
%load_ext ipythontutor
```

Les opérateurs `is` et `==` ~~~~~

- nous avons déjà parlé de l'opérateur `==` qui **compare la valeur** de deux objets ;
- python fournit aussi un opérateur `is` qui permet de savoir si deux valeurs correspondent **au même objet** en mémoire.

Nous allons illustrer la différence entre ces deux opérateurs.

```
# deux listes identiques
a = [1, 2]
b = [1, 2]

# les deux objets se ressemblent
print('==', a == b)
```

```
== True
```

```
# mais ce ne sont pas les mêmes objets
print('is', a is b)
```

```
is False
```

```
# par contre ici il n'y a qu'une liste
a = [1, 2]

# et les deux variables pointent vers le même objet
b = a

# non seulement les deux expressions se ressemblent
print('==', a == b)
```

```
== True
```

```
# mais elles désignent le même objet
print('is', a is b)
```

```
is True
```

60.1.1 La même chose sous pythontutor

```
%%ipythontutor curInstr=2
a = [1, 2]
b = [1, 2]
```

```
%%ipythontutor curInstr=1
# équivalent à la forme ci-dessus
a = b = [1, 2]
```

60.1.2 Utilisez `is` plutôt que `==` lorsque c'est possible

La pratique usuelle est d'utiliser `is` lorsqu'on compare avec un objet qui est un singleton, comme typiquement `None`.

Par exemple on préférera écrire :

```
undef = None

if undef is None:
    print('indéfini')
```

```
indéfini
```

plutôt que

```
if undef == None:
    print('indéfini')
```

```
indéfini
```

qui se comporte de la même manière (à nouveau, parce qu'on compare avec `None`), mais est légèrement moins lisible, et franchement moins pythonique :)

Notez aussi et surtout que `is` est **plus efficace** que `==`. En effet `is` peut être évalué en temps constant, puisqu'il s'agit essentiellement de comparer les deux adresses. Alors que pour `==` il peut s'agir de parcourir toute une structure de données possiblement très complexe.

60.2 Complément - niveau intermédiaire

60.2.1 La fonction `id`

Pour bien comprendre le fonctionnement de `is` nous allons voir la fonction `id` qui retourne un identificateur unique pour chaque objet ; un modèle mental acceptable est celui d'adresse mémoire.

```
id(True)
```



```
140414904648224
```

Comme vous vous en doutez, l'opérateur `is` peut être décrit formellement à partir de `id` comme ceci

$$(a \text{ is } b) \iff (\text{id}(a) == \text{id}(b))$$

60.2.2 Certains types de base sont des singletons

Un singleton est un objet qui n'existe qu'en un seul exemplaire dans la mémoire. Un usage classique des singletons en python est de minimiser le nombre d'objets immuables en mémoire. Voyons ce que cela nous donne avec des entiers

```
a = 3
b = 3
print('a', id(a), 'b', id(b))
```

```
a 140414905057088 b 140414905057088
```

Tiens, c'est curieux, nous avons ici deux objets, que l'on pourrait penser différents, mais en fait ce sont les mêmes ; `a` et `b` désignent **le même objet** python, et on a

```
a is b
```

```
True
```

Il se trouve que, dans le cas des petits entiers, python réalise une optimisation de l'utilisation de la mémoire. Quel que soit le nombre de variables dont la valeur est 3, un seul objet correspondant à l'entier 3 est alloué et créé, pour éviter d'engorger la mémoire. On dit que l'entier 3 est implémenté comme un singleton ; nous reverrons ceci en exercice.

On trouve cette optimisation avec quelques autres objets python, comme par exemple

```
a = ""
b = ""
a is b
```

```
True
```

Ou encore, plus surprenant :

```
a = "foo"
b = "foo"
a is b
```

```
True
```

Conclusion cette optimisation ne touche aucun type mutable (heureusement) ; pour les types immuables, il n'est pas extrêmement important de savoir en détail quels objets sont implémentés de la sorte.

Ce qui est par contre extrêmement important est de comprendre la différence entre `is` et `==`, et de les utiliser à bon escient au risque d'écrire du code fragile.

60.2.3 Pour en savoir plus

Aux étudiants de niveau avancé, nous recommandons la lecture de la section « Objects, values and types » dans la documentation python

<https://docs.python.org/3/reference/datamodel.html#objects-values-and-types>

qui aborde également la notion de « garbage collection », que nous n’aurons pas le temps d’approfondir dans ce MOOC.

Listes infinies & références circulaires

61.1 Complément - niveau intermédiaire

```
%load_ext ipythontutor
```

Nous allons maintenant construire un objet un peu abscons. Cet exemple précis n'a aucune utilité pratique, mais permet de bien comprendre la logique du langage.

Construisons une liste à un seul élément, peu importe quoi :

```
infini_1 = [None]
```

À présent nous allons remplacer le premier et seul élément de la liste par... la liste elle-même

```
infini_1[0] = infini_1  
print(infini_1)
```

```
[[...]]
```

Pour essayer de décrire l'objet liste ainsi obtenu, on pourrait dire qu'il s'agit d'une liste de taille 1 et de profondeur infinie, une sorte de fil infini en quelque sorte.

Naturellement, l'objet obtenu est difficile à imprimer de manière convaincante. Pour faire en sorte que cet objet soit tout de même imprimable, et éviter une boucle infinie, python utilise l'ellipse ... pour indiquer ce qu'on appelle une référence circulaire. Si on n'y prenait pas garde en effet, il faudrait écrire [[[[[etc.]]]] avec une infinité de crochets.

Voici la même séquence exécutée sous <http://pythontutor.com> ; il s'agit d'un site très utile pour comprendre comment python implémente les objets, les références et les partages.

Cliquez sur le bouton `Forward` pour avancer dans l'exécution de la séquence. À la fin de la séquence vous verrez - ce n'est pas forcément clair - la seule cellule de la liste à se référencer elle-même :

```
%%ipythontutor height=230  
infini_1 = [None]  
infini_1[0] = infini_1
```

Toutes les fonctions de python ne sont pas aussi intelligentes que `print`. Bien qu'on puisse comparer cette liste avec elle-même :

```
infini_1 == infini_1
```

```
True
```

il n'en est pas de même si on la compare avec un objet analogue mais pas identique :

```
infini_2 = [0]
infini_2[0] = infini_2
print(infini_2)
infini_1 == infini_2
```

```
[[...]]
```

```
-----
RecursionError                                Traceback (most recent call last)
<ipython-input-4-d9e3869156cb> in <module>()
      2 infini_2[0] = infini_2
      3 print(infini_2)
----> 4 infini_1 == infini_2

RecursionError: maximum recursion depth exceeded in comparison
```

61.1.1 Généralisation aux références circulaires

On obtient un phénomène équivalent dès lors qu'un élément contenu dans un objet fait référence à l'objet lui-même. Voici par exemple comment on peut construire un dictionnaire qui contient une référence circulaire :

```
collection_de_points = [
    {'x': 10, 'y': 20},
    {'x': 30, 'y': 50},
    # imaginez plein de points
]

# on rajoute dans chaque dictionnaire une clé 'points'
# qui référence la collection complète
for point in collection_de_points:
    point['points'] = collection_de_points

# la structure possède maintenant des références circulaires
print(collection_de_points)
```

```
[{'x': 10, 'y': 20, 'points': [...]}, {'x': 30, 'y': 50, 'points': [...]}]
```

On voit à nouveau réapparaître les éllipses, qui indiquent que pour chaque point, le nouveau champ “points” est un objet qui a déjà été imprimé.

Cette technique est cette fois très utile et très utilisée dans la pratique, dès lors qu'on a besoin de naviguer de manière arbitraire dans une structure de données compliquée. Dans cet exemple, pas très réaliste naturellement, on pourrait à présent accéder depuis un point à tous les autres points de la collection dont il fait partie.

À nouveau il peut être intéressant de voir le comportement de cet exemple avec <http://pythontutor.com> pour bien comprendre ce qui se passe, si cela ne vous semble pas clair à première vue :

```
%%ipythontutor curInstr=7
points = [
    {'x': 10, 'y': 20},
```

```
{'x': 30, 'y': 50},  
]  
  
for point in points:  
    point['points'] = points
```

Les différentes copies

```
%load_ext ipythontutor
```

62.1 Complément - niveau basique

62.1.1 Deux types de copie

Pour résumer les deux grands types de copie que l'on a vues dans la vidéo :

- La *shallow copy* - de l'anglais *shallow* qui signifie superficiel
- La *deep copy* - de *deep* qui signifie profond

62.1.2 Le module `copy`

Pour réaliser une copie, la méthode la plus simple, en ceci qu'elle fonctionne avec tous les types de manière identique, consiste à utiliser *le module standard* '`copy`' <<https://docs.python.org/3/library/copy.html>>, et notamment

- `copy.copy` pour une copie superficielle
- `copy.deepcopy` pour une copie en profondeur

```
import copy
#help(copy.copy)
#help(copy.deepcopy)
```

62.1.3 Un exemple

Nous allons voir le résultat des deux formes de copies sur un même sujet de départ.

La copie superficielle / *shallow* copie / `copy.copy`

N'oubliez pas de cliquer le bouton Forward dans la fenêtre pythontutor :

```
%%ipythontutor height=410 curInstr=6
import copy
# On se donne un objet de départ
source = [
    [1, 2, 3], # une liste
    {1, 2, 3}, # un ensemble
    (1, 2, 3), # un tuple
    '123',     # un string
    123,       # un entier
]
# une copie simple renvoie ceci
shallow_copy = copy.copy(source)
```

Vous remarquez que

- la source et la copie partagent tous leurs (sous-)éléments, et notamment la liste `source[0]` et l'ensemble `source[1]` ;
- ainsi, après cette copie, on peut modifier l'un de ces deux objets (la liste ou l'ensemble), et ainsi modifier la source **et** la copie ;

On rappelle aussi que, la source étant une liste, on aurait pu aussi bien faire la copie superficielle avec

```
shallow2 = source[:]
```

La copie profonde / *deep* copie / `copy.deepcopy`

Sur le même objet de départ, voici ce que fait la copie profonde :

```
%%ipythontutor height=410 curInstr=6
import copy
# On se donne un objet de départ
source = [
    [1, 2, 3], # une liste
    {1, 2, 3}, # un ensemble
    (1, 2, 3), # un tuple
    '123',     # un string
    123,       # un entier
]
# une copie profonde renvoie ceci
deep_copy = copy.deepcopy(source)
```

Ici, il faut remarquer que

- les deux objets mutables accessibles via `source`, c'est-à-dire **la liste** `source[0]` et **l'ensemble** `source[1]`, ont été tous deux dupliqués ;
- **le tuple** correspondant à `source[2]` n'est **pas dupliqué**, mais comme il n'est **pas mutable** on ne peut pas modifier la copie au travers de la source ;
- de manière générale, on a la bonne propriété que la source et sa copie ne partagent rien qui soit modifiable,
- et donc on ne peut pas modifier l'un au travers de l'autre.

On retrouve donc à nouveau l'optimisation qui est mise en place dans python pour implémenter les types immuables comme des singletons lorsque c'est possible. Cela a été vu en détail dans le complément consacré à l'opérateur `is`.

62.2 Complément - niveau intermédiaire

```
# on répète car le code précédent a seulement été exposé à pythontutor
import copy
source = [
    [1, 2, 3], # une liste
    {1, 2, 3}, # un ensemble
    (1, 2, 3), # un tuple
    '123',     # un string
    123,       # un entier
]
shallow_copy = copy.copy(source)
deep_copy = copy.deepcopy(source)
```

62.2.1 Objets égaux au sens logique

Bien sûr ces trois objets se ressemblent si on fait une comparaison *logique* avec ==

```
print('source == shallow_copy:', source == shallow_copy)
print('source == deep_copy:', source == deep_copy)
```

```
source == shallow_copy: True
source == deep_copy: True
```

62.2.2 Inspectons les objets de premier niveau

Mais par contre si on compare l'**identité** des objets de premier niveau, on voit que source et shallow_copy partagent leurs objets :

```
# voir la cellule ci-dessous si ceci vous parait peu clair
for i, (source_item, copy_item) in enumerate(zip(source, shallow_copy)):
    compare = source_item is copy_item
    print(f"source[{i}] is shallow_copy[{i}] -> {compare}")
```

```
source[0] is shallow_copy[0] -> True
source[1] is shallow_copy[1] -> True
source[2] is shallow_copy[2] -> True
source[3] is shallow_copy[3] -> True
source[4] is shallow_copy[4] -> True
```

```
# rappel au sujet de zip et enumerate
# la cellule ci-dessous est essentiellement équivalente à
for i in range(len(source)):
    compare = source[i] is shallow_copy[i]
    print(f"source[{i}] is shallow_copy[{i}] -> {compare}")
```

```
source[0] is shallow_copy[0] -> True
source[1] is shallow_copy[1] -> True
source[2] is shallow_copy[2] -> True
source[3] is shallow_copy[3] -> True
source[4] is shallow_copy[4] -> True
```

Alors que naturellement ce **n'est pas le cas** avec la copie en profondeur

```
# voir la cellule ci-dessous si ceci vous parait peu clair
for i, (source_item, deep_item) in enumerate(zip(source, deep_copy)):
    compare = source_item is deep_item
    print(f"source[{i}] is deep_copy[{i}] -> {compare}")
```

```
compare = source_item is deep_item
print(f"source[{i}] is deep_copy[{i}] -> {compare}")
```

```
source[0] is deep_copy[0] -> False
source[1] is deep_copy[1] -> False
source[2] is deep_copy[2] -> True
source[3] is deep_copy[3] -> True
source[4] is deep_copy[4] -> True
```

On retrouve ici ce qu'on avait déjà remarqué sous pythontutor, à savoir que les trois derniers objets - immutables - n'ont pas été dupliqués comme on aurait pu s'y attendre.

62.2.3 On modifie la source

Il doit être clair à présent que, précisément parce que `deep_copy` est une copie en profondeur, on peut modifier `source` sans impacter du tout `deep_copy`.

S'agissant de `shallow_copy`, par contre, seuls les éléments de premier niveau ont été copiés. Aussi si on fait une modification par exemple à l'intérieur de la liste qui est le premier fils de `source`, cela sera **répercuté** dans `shallow_copy`

```
print("avant, source", source)
print("avant, shallow_copy", shallow_copy)
source[0].append(4)
print("après, source", source)
print("après, shallow_copy", shallow_copy)
```

```
avant, source      [[1, 2, 3], {1, 2, 3}, (1, 2, 3), '123', 123]
avant, shallow_copy [[1, 2, 3], {1, 2, 3}, (1, 2, 3), '123', 123]
après, source      [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
après, shallow_copy [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
```

Si par contre on remplace complètement un élément de premier niveau dans la source, cela ne sera pas répercuté dans la copie superficielle

```
print("avant, source", source)
print("avant, shallow_copy", shallow_copy)
source[0] = 'remplacement'
print("après, source", source)
print("après, shallow_copy", shallow_copy)
```

```
avant, source      [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
avant, shallow_copy [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
après, source      ['remplacement', {1, 2, 3}, (1, 2, 3), '123', 123]
après, shallow_copy [[1, 2, 3, 4], {1, 2, 3}, (1, 2, 3), '123', 123]
```

62.2.4 Copie et circularité

Le module `copy` est capable de copier - même en profondeur - des objets contenant des références circulaires.

```
l = [None]
l[0] = l
l
```

```
[[...]]
```

```
copy.copy(l)
```

```
[[[...]]]
```

```
copy.deepcopy(l)
```

```
[...] ]]
```

62.2.5 Pour en savoir plus

On peut se reporter à la section sur le module ‘*copy*’ <<https://docs.python.org/3/library/copy.html>> __ dans la documentation python.

63.1 Complément - niveau basique

Voici un récapitulatif sur l'instruction `del` selon le contexte dans lequel elle est utilisée.

63.1.1 Sur une variable

On peut annuler la définition d'une variable. Pour anticiper sur les exceptions, nous attrapons ici l'exception `NameError` qui est produite lorsqu'une variable est utilisée alors qu'elle n'est pas définie. Nous discuterons du fonctionnement des exceptions cette semaine.

```
# la variable a n'est pas définie
try:
    print('a=', a)
except NameError as e:
    print("a n'est pas définie")
```

```
a n'est pas définie
```

```
# on la définit
a = 10

# aucun souci ici, l'exception n'est pas levée
try:
    print('a=', a)
except NameError as e:
    print("a n'est pas définie")
```

```
a= 10
```

```
# maintenant on peut effacer la variable
del a

# c'est comme si on ne l'avait pas définie
# dans la cellule précédente
try:
```

```
print('a=', a)
except NameError as e:
    print("a n'est pas définie")
```

```
a n'est pas définie
```

63.1.2 Sur une liste

On peut enlever d'une liste les éléments qui correspondent à une *slice* :

```
# on se donne une liste
l = list(range(12))
print(l)
```

```
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
```

```
# on considère une slice dans cette liste
print('slice=', l[2:10:3])

# voyons ce que ça donne si on efface cette slice
del l[2:10:3]
print("après del", l)
```

```
slice= [2, 5, 8]
après del [0, 1, 3, 4, 6, 7, 9, 10, 11]
```

63.1.3 Sur un dictionnaire

Avec `del` on peut enlever une clé, et donc la valeur correspondante, d'un dictionnaire :

```
# partons d'un dictionnaire simple
d = dict(foo='bar', spam='eggs', a='b')
print(d)
```

```
{'foo': 'bar', 'spam': 'eggs', 'a': 'b'}
```

```
# on peut enlever une clé avec del
del d['a']
print(d)
```

```
{'foo': 'bar', 'spam': 'eggs'}
```

63.1.4 On peut passer plusieurs arguments à `del`

```
# Voyons où en sont nos données
print('l', l)
print('d', d)
```

```
l [0, 1, 3, 4, 6, 7, 9, 10, 11]
d {'foo': 'bar', 'spam': 'eggs'}
```

```
# on peut invoquer 'del' avec plusieurs expressions
# séparées par une virgule

del l[3:], d['spam']

print('l', l)
print('d', d)
```

```
l [0, 1, 3]
d {'foo': 'bar'}
```

63.1.5 Pour en savoir plus

La page sur l'instruction `'del'` <https://docs.python.org/3/reference/simple_stmts.html#the-del-statement> dans la documentation python

Affectation simultanée

64.1 Complément - niveau basique

Nous avons déjà parlé de l'affectation par *sequence unpacking* (en Semaine 3, séquence « Les tuples »), qui consiste à affecter à plusieurs variables des « morceaux » d'un objet, comme dans :

```
x, y = ['spam', 'egg']
```

Dans ce complément nous allons voir une autre forme de l'affectation, qui consiste à affecter **le même objet** à plusieurs variables. Commençons par un exemple simple :

```
a = b = 1
print('a', a, 'b', b)
```

```
a 1 b 1
```

La raison pour laquelle nous abordons cette construction maintenant est qu'elle a une forte relation avec les références partagées ; pour bien le voir, nous allons utiliser une valeur mutable comme valeur à affecter :

```
# on affecte a et b au même objet liste vide
a = b = []
```

Dès lors nous sommes dans le cas typique d'une référence partagée ; une modification de `a` va se répercuter sur `b` puisque ces deux variables désignent **le même objet** :

```
a.append(1)
print('a', a, 'b', b)
```

```
a [1] b [1]
```

Ceci est à mettre en contraste avec plusieurs affectations séparées :

```
# si on utilise deux affectations différentes
a = []
b = []

# alors on peut changer a sans changer b
```

```
a.append(1)
print('a', a, 'b', b)
```

```
a [1] b []
```

On voit que dans ce cas chaque affectation crée une liste vide différente, et les deux variables ne partagent plus de donnée.

D'une manière générale, utiliser l'affectation simultanée vers un objet mutable crée mécaniquement des **références partagées**, aussi vérifiez bien dans ce cas que c'est votre intention.

Les instructions += et autres revisit  es

65.1 Compl  ment - niveau interm  diaire

Nous avons vu en premi  re semaine (S  quence « Les types num  riques ») une premi  re introduction aux instructions += et ses d  riv  es comme *=, **=, etc.

65.1.1 Ces constructions ont une d  finition    g  om  trie variable

En C quand on utilise += (ou encore ++) on modifie la m  moire en place - historiquement, cet op  rateur permettait au programmeur d'aider    l'optimisation du code pour utiliser les instructions assembleur idoines.

Ces constructions en python s'inspirent clairement de C, aussi dans l'esprit ces constructions devraient fonctionner en **modifiant** l'objet r  f  renc   par la variable.

Mais les types num  riques en python ne sont **pas mutables**, alors que les listes le sont. Du coup le comportement de += est **diff  rent** selon qu'on l'utilise sur un nombre ou sur une liste, ou plus g  n  ralement selon qu'on l'invoque sur un type mutable ou non. Voyons cela sur des exemples tr  s simples.

```
# Premier exemple avec un entier

# on commence avec une r  f  rence partag  e
a = b = 3
a is b
```

```
True
```

```
# on utilise += sur une des deux variables
a += 1

# ceci n'a pas modifi   b
# c'est normal, l'entier n'est pas mutable

print(a)
print(b)
print(a is b)
```

```
4
3
False
```

```
# Deuxième exemple, cette fois avec une liste

# la même référence partagée
a = b = []
a is b
```

```
True
```

```
# pareil, on fait += sur une des variables
a += [1]

# cette fois on a modifié a et b
# car += a pu modifier la liste en place
print(a)
print(b)
print(a is b)
```

```
[1]
[1]
True
```

Vous voyez donc que la sémantique de += (c'est bien entendu le cas pour toutes les autres formes d'instructions qui combinent l'affectation avec un opérateur) **est différente** suivant que l'objet référencé par le terme de gauche est **mutable ou immuable**.

Pour cette raison, c'est là une opinion personnelle, cette famille d'instructions n'est pas le trait le plus réussi dans le langage, et je ne recommande pas de l'utiliser.

65.1.2 Précision sur la définition de +=

Nous avons dit en première semaine, et en première approximation, que

```
x += y
```

était équivalent à

```
x = x + y
```

Au vu de ce qui précède, on voit que ce n'est **pas tout à fait exact**, puisque :

```
# si on fait x += y sur une liste
# on fait un effet de bord sur la liste
# comme on vient de le voir

a = []
print("avant", id(a))
a += [1]
print("après", id(a))
```

```
avant 140354234518152
après 140354234518152
```

```
# alors que si on fait x = x + y sur une liste
# on crée un nouvel objet liste
```

```
a = []  
print("avant", id(a))  
a = a + [1]  
print("après", id(a))
```

```
avant 140354234518408  
après 140354234517960
```

Vous voyez donc que vis-à-vis des références partagées, ces deux façons de faire mènent à un résultat différent.

66.1 Exercice - niveau basique

```
# charger l'exercice
from corrections.cls_fifo import exo_fifo
```

On veut implémenter une classe pour manipuler une queue d'événements. La logique de cette classe est que :

- on la crée sans argument,
- on peut toujours ajouter un élément avec la méthode `incoming`;
- et tant que la queue contient des éléments on peut appeler la méthode `outgoing`, qui retourne et enlève un élément dans la queue.

Cette classe s'appelle `Fifo` pour *First in, first out*, c'est-à-dire que les éléments retournés par `outgoing` le sont dans le même ordre où ils ont été ajoutés.

La méthode `outgoing` retourne `None` lorsqu'on l'appelle sur une pile vide.

```
# voici un exemple de scénario
exo_fifo.example()
```

```
# vous pouvez définir votre classe ici

class Fifo:
    def __init__(self):
        "votre code"
    def incoming(self, value):
        "votre code"
    def outgoing(self):
        "votre code"
```

```
# et la vérifier ici
exo_fifo.correction(Fifo)
```

Passage d'arguments par référence

67.1 Complément - niveau intermédiaire

Entre le code qui appelle une fonction et le code de la fonction elle-même

```
def ma_fonction(dans_fonction):  
    print(dans_fonction)  
  
dans_appelant = ["texte"]  
ma_fonction(dans_appelant)
```

```
['texte']
```

on peut se demander quelle est exactement la nature de la relation entre l'appelant et l'appelé, c'est-à-dire ici `dans_appelant` et `dans_fonction`.

C'est l'objet de ce complément.

67.1.1 Passage par valeur - passage par référence

Si vous avez appris d'autres langages de programmation comme C ou C++, on a pu vous parler de deux modes de passage de paramètres :

- par valeur : cela signifie qu'on communique à la fonction, non pas l'entité dans l'appelant, mais seulement **sa valeur**; en clair, **une copie** ;
- par référence : cela signifie qu'on passe à la fonction une **référence** à l'argument dans l'appelant, donc essentiellement les deux codes **partagent** la même mémoire.

67.1.2 Python fait du passage par référence

Certains langages comme Pascal - et C++ si on veut - proposent ces deux modes. En python, tous les passages de paramètres se font **par référence**.

```
# chargeons la magie pour pythontutor  
%load_ext ipythontutor
```

```
%%ipythontutor curInstr=4
def ma_fonction(dans_fonction):
    print(dans_fonction)

dans_appelant = ["texte"]
ma_fonction(dans_appelant)
```

Ce qui signifie qu'on peut voir le code ci-dessus comme étant - pour simplifier - équivalent à ceci :

```
dans_appelant = ["texte"]

# ma_fonction (dans_appelant)
# → on entre dans la fonction
dans_fonction = dans_appelant
print(dans_fonction)
```

```
['texte']
```

On peut le voir encore d'une autre façon en instrumentant le code comme ceci – on rappelle que la fonction built-in `id` retourne l'adresse mémoire d'un objet :

```
def ma_fonction(dans_fonction):
    print('dans ma_fonction', dans_fonction , id(dans_fonction))

dans_appelant = ["texte"]
print('dans appelant ', dans_appelant, id(dans_appelant))
ma_fonction(dans_appelant)
```

```
dans appelant      ['texte'] 139847418555784
dans ma_fonction  ['texte'] 139847418555784
```

67.1.3 Des références partagées

On voit donc que l'appel de fonction crée des références partagées, exactement comme l'affectation, et que tout ce que nous avons vu au sujet des références partagées s'applique exactement à l'identique :

```
# on ne peut pas modifier un immuable dans une fonction
def increment(n):
    n += 1

compteur = 10
increment(compteur)
print(compteur)
```

```
10
```

```
# on peut par contre ajouter dans une liste
def insert(liste, valeur):
    liste.append(valeur)

liste = ["un"]
insert(liste, "texte")
print(liste)
```

```
['un', 'texte']
```

Pour cette raison, il est important de bien préciser, quand vous documentez une fonction, si elle fait des effets de bord sur ses arguments (c'est-à-dire qu'elle modifie ses arguments), ou si elle produit une copie. Rappelez-vous

par exemple le cas de la méthode `sort` sur les listes, et de la fonction de commodité `sorted`, que nous avons vues en semaine 2.

De cette façon, on saura s'il faut ou non copier l'argument avant de le passer à votre fonction.

Rappels sur *docstring*

68.1 Complément - niveau basique

68.1.1 Comment documenter une fonction

Pour rappel, il est recommandé de toujours documenter les fonctions en ajoutant une chaîne comme première instruction.

```
def flatten(containers):  
    "returns a list of the elements of the elements in containers"  
    return [element for container in containers for element in container]
```

Cette information peut être consultée, soit interactivement :

```
help(flatten)
```

```
Help on function flatten in module __main__:
```

```
flatten(containers)  
    returns a list of the elements of the elements in containers
```

Soit programmativement :

```
flatten.__doc__
```

```
'returns a list of the elements of the elements in containers'
```

68.1.2 Sous quel format ?

L'usage est d'utiliser une chaîne simple (délimitée par « " » ou « ' ») lorsque le *docstring* tient sur une seule ligne, comme ci-dessus.

Lorsque ce n'est pas le cas - et pour du vrai code, c'est rarement le cas - on utilise des chaînes multi-lignes (délimitées par « """ » ou « ''' »). Dans ce cas le format est très flexible, car le *docstring* est normalisé, comme on le voit sur ces deux exemples, où le rendu final est identique :

```
# un style de docstring multi-lignes
def flatten(containers):
    """
    provided that containers is a list (or more generally an iterable)
    of elements that are themselves iterables, this function
    returns a list of the items in these elements
    """
    return [element for container in containers for element in container]

help(flatten)
```

```
Help on function flatten in module __main__:

flatten(containers)
    provided that containers is a list (or more generally an iterable)
    of elements that are themselves iterables, this function
    returns a list of the items in these elements
```

```
# un autre style, qui donne le même résultat
def flatten(containers):
    """
    provided that containers is a list (or more generally an iterable)
    of elements that are themselves iterables, this function
    returns a list of the items in these elements
    """
    return [element for container in containers for element in container]

help(flatten)
```

```
Help on function flatten in module __main__:

flatten(containers)
    provided that containers is a list (or more generally an iterable)
    of elements that are themselves iterables, this function
    returns a list of the items in these elements
```

68.1.3 Quelle information ?

On remarquera que dans ces exemples, le *docstring* ne répète pas le nom de la fonction ou des arguments (en mots savants, sa *signature*), et que ça n’empêche pas `help` de nous afficher cette information.

Le [PEP 257](#) qui donne les conventions autour du *docstring* précise bien ceci :

The one-line docstring should NOT be a “signature” reiterating the function/method parameters (which can be obtained by introspection). Don’t do :

```
def function(a, b):
    """function(a, b) -> list"""
```

<...>

The preferred form for such a docstring would be something like :

```
def function(a, b):
    """Do X and return a list."""
```

(Of course “Do X” should be replaced by a useful description !)

68.1.4 Pour en savoir plus

Vous trouverez tous les détails sur *docstring* dans le [PEP 257](#).

69.1 Complément - niveau basique

69.1.1 Typage dynamique

En première semaine, nous avons rapidement mentionné les concepts de typage statique et dynamique.

Avec la fonction prédéfinie `isinstance` - qui peut être par ailleurs utile dans d'autres contextes - vous pouvez facilement :

- vérifier qu'un argument d'une fonction a bien le type attendu,
- et traiter différemment les entrées selon leur type.

Voyons tout de suite sur un exemple simple comment on pourrait définir une fonction qui travaille sur un entier, mais qui par commodité peut aussi accepter un entier passé comme une chaîne de caractères, ou même une liste d'entiers (auquel cas on renvoie la liste des factorielles) :

```
def factoriel(argument):
    # si on reçoit un entier
    if isinstance(argument, int):
        # (*)
        return 1 if argument <= 1 else argument * factoriel(argument - 1)
    # convertir en entier si on reçoit une chaîne
    elif isinstance(argument, str):
        return factoriel(int(argument))
    # la liste des résultats si on reçoit un tuple ou une liste
    elif isinstance(argument, (tuple, list)): # (**)
        return [factoriel(i) for i in argument]
    # sinon on lève une exception
    else:
        raise TypeError(argument)
```

```
print("entier", factoriel(4))
print("chaîne", factoriel("8"))
print("tuple", factoriel((4, 8)))
```

```
entier 24
chaîne 40320
tuple [24, 40320]
```

Remarquez que la fonction `isinstance` **possède elle-même** une logique de ce genre, puisqu'en ligne 3 (*) nous lui avons passé en deuxième argument un type (`int`), alors qu'en ligne 11 (**) on lui a passé un tuple de deux types. Dans ce second cas naturellement, elle vérifie si l'objet (le premier argument) est **de l'un des types** mentionnés dans le tuple.

69.2 Complément - niveau intermédiaire

69.2.1 Le module `types`

Le module `types` définit un certain nombre de constantes qui peuvent être utiles dans ce contexte - vous trouverez une liste exhaustive à la fin de ce notebook. Par exemple :

```
from types import FunctionType
isinstance(factoriel, FunctionType)
```

```
True
```

Mais méfiez vous toutefois des fonctions *built-in*, qui sont de type `BuiltinFunctionType`

```
from types import BuiltinFunctionType
isinstance(len, BuiltinFunctionType)
```

```
True
```

```
# alors qu'on pourrait penser que
isinstance(len, FunctionType)
```

```
False
```

69.2.2 `isinstance` vs `type`

Il est recommandé d'utiliser `isinstance` par rapport à la fonction `type`. Tout d'abord, cela permet, on vient de le voir, de prendre en compte plusieurs types.

Mais aussi et surtout `isinstance` supporte la notion d'héritage qui est centrale dans le cadre de la programmation orientée objet, sur laquelle nous allons anticiper un tout petit peu par rapport aux présentations de la semaine prochaine.

Avec la programmation objet, vous pouvez définir vos propres types. On peut par exemple définir une classe `Animal` qui convient pour tous les animaux, puis définir une sous-classe `Mammifere`. On dit que la classe `Mammifere` *hérite* de la classe `Animal`, et on l'appelle sous-classe parce qu'elle représente une partie des animaux ; et donc tout ce qu'on peut faire sur les animaux peut être fait sur les mammifères.

En voici une implémentation très rudimentaire, uniquement pour illustrer le principe de l'héritage. Si ce qui suit vous semble difficile à comprendre, pas d'inquiétude, nous reviendrons sur ce sujet lorsque nous parlerons des classes.

```
class Animal:
    def __init__(self, name):
        self.name = name

class Mammifere(Animal):
```

```
def __init__(self, name):
    Animal.__init__(self, name)
```

Ce qui nous intéresse dans l’immédiat c’est que `isinstance` permet dans ce contexte de faire des choses qu’on ne peut pas faire directement avec la fonction `type`, comme ceci :

```
# c'est comme ceci qu'on peut créer un objet de type `Animal` (méthode __init__)
requin = Animal('requin')
# idem pour un Mammifere
baleine = Mammifere('baleine')

# bien sûr ici la réponse est 'True'
print("l'objet baleine est-il un mammifere ?", isinstance(baleine, Mammifere))
```

```
l'objet baleine est-il un mammifere ? True
```

```
# ici c'est moins évident, mais la réponse est 'True' aussi
print("l'objet baleine est-il un animal ?", isinstance(baleine, Animal))
```

```
l'objet baleine est-il un animal ? True
```

Vous voyez qu’ici, bien que l’objet `baleine` est de type `Mammifere`, on peut le considérer comme étant **aussi** de type `Animal`.

Ceci est motivé de la façon suivante : comme on l’a dit plus haut, tout ce qu’on peut faire (en termes notamment d’envoi de méthodes) sur un objet de type `Animal`, on peut le faire sur un objet de type `Mammifere`. Dit en termes ensemblistes, l’ensemble des mammifères est inclus dans l’ensemble des animaux.

69.2.3 Annexe - Les symboles du module `types`

Vous pouvez consulter la documentation du modules ‘types’ <<https://docs.python.org/3/library/types.html>> ‘__.

```
# voici par ailleurs la liste de ses attributs
import types
dir(types)
```

```
['AsyncGeneratorType',
 'BuiltinFunctionType',
 'BuiltinMethodType',
 'CodeType',
 'CoroutineType',
 'DynamicClassAttribute',
 'FrameType',
 'FunctionType',
 'GeneratorType',
 'GetSetDescriptorType',
 'LambdaType',
 'MappingProxyType',
 'MemberDescriptorType',
 'MethodType',
 'ModuleType',
 'SimpleNamespace',
 'TracebackType',
 '_GeneratorWrapper',
 '__all__',
 '__builtins__',
 '__cached__',
 '__doc__',
 '__file__',
 '__loader__',
```

```
'__name__',  
'__package__',  
'__spec__',  
'_ag',  
'_calculate_meta',  
'_collections_abc',  
'_functools',  
'coroutine',  
'new_class',  
'prepare_class']
```

70.1 Complément - niveau intermédiaire

70.1.1 Langages compilés

Nous avons évoqué en première semaine le typage, lorsque nous avons comparé python avec les langages compilés. Dans un langage compilé avec typage statique, on **doit fournir du typage**, ce qui fait qu'on écrit typiquement une fonction comme ceci :

```
int factoriel(int n) {  
    return (n<=1) ? 1 : n * factoriel(n-1);  
}
```

ce qui signifie que la fonction factoriel prend un premier argument qui est un entier, et qu'elle retourne également un entier.

Nous avons vu également que, par contraste, pour écrire une fonction en python, on n'a **pas besoin** de préciser le **type** des arguments ni du retour de la fonction.

70.1.2 Vous pouvez aussi typer votre code python

Cependant depuis la version 3.5, python supporte un mécanisme **totalement optionnel** qui vous permet d'annoter les arguments des fonctions avec des informations de typage, ce mécanisme est connu sous le nom de *type hints*, et ça se présente comme ceci.

typer une variable

```
# pour typer une variable avec les type hints  
nb_items : int = 0
```

```
nb_items
```

```
0
```

typer les paramètres et le retour d'une fonction

```
# une fonction factorielle avec des type hints
def fact(n : int) -> int:
    return 1 if n <= 1 else n * fact(n-1)
```

```
fact(12)
```

```
479001600
```

70.1.3 Usages

À ce stade, on peut entrevoir les usages suivants à ce type d'annotation :

- tout d'abord, et évidemment, cela peut permettre de mieux documenter le code ;
- les environnements de développement sont susceptibles de vous aider de manière plus effective ; si à quelque part vous écrivez `z = fact(12)`, le fait de savoir que `z` est entier permet de fournir une complétion plus pertinente lorsque vous commencez à écrire `z . [TAB]` ;
- on peut espérer trouver des erreurs dans les passages d'arguments à un stade plus précoce du développement.

Par contre ce qui est très très clairement annoncé également, c'est que ces informations de typage sont **totale**ment **facultatives**, et que le langage les **ignore totalement**.

```
# l'interpréteur ignore totalement ces informations
def fake_fact(n : str) -> str:
    return 1 if n <= 1 else n * fake_fact(n-1)

# on peut appeler fake_fact avec un int alors
# que c'est déclaré pour des str
fake_fact(12)
```

```
479001600
```

Le modèle préconisé est d'utiliser des **outils extérieurs**, qui peuvent faire une analyse statique du code pour exploiter ces informations à des fins de validation. Dans cette catégorie, le plus célèbre *est sans doute* 'mypy' <<http://mypy-lang.org/>>__. Notez aussi que les IDE comme PyCharm sont également capables de tirer parti de ces annotations.

70.1.4 Est-ce répandu ?

Parce qu'ils ont été introduits pour la première fois avec python-3.5, en 2015 donc, puis améliorés dans la 3.6 pour le typage des variables, l'usage des *type hints* n'est pour l'instant pas très répandu, en proportion de code en tous cas. En outre, il aura fallu un temps de latence avant que tous les outils (IDE's, producteurs de documentation, outils de test, validateurs...) ne soient améliorés pour en tirer un profit maximal.

On peut penser que cet usage va se répandre avec le temps, peut-être / sans doute pas de manière systématique, mais *a minima* pour lever certaines ambiguïtés.

70.1.5 Comment annoter son code

Maintenant que nous en avons bien vu la finalité, voyons un très bref aperçu des possibilités offertes pour la construction des types dans ce contexte de *type hints*. N'hésitez pas à vous reporter à la documentation officielle du module 'typing' <<https://docs.python.org/3/library/typing.html>>__ pour un exposé plus exhaustif.

le module `typing`

L'ensemble des symboles que nous allons utiliser dans la suite de ce complément provient du module `typing`

exemples simples

```
from typing import List
```

```
# une fonction qui
# attend à un paramètre qui soit une liste d'entiers,
# et qui retourne une liste de chaines
def foo(x: List[int]) -> List[str]:
    pass
```

avertissement : `list` vs `List`

Remarquez bien dans l'exemple ci-dessus que nous avons utilisé `typing.List` plutôt que le type builtin `list`, alors que l'on a pu par contre utiliser `int` et `str`.

Les raisons pour cela sont de deux ordres :

- tout d'abord, si je devais utiliser `list` pour construire un type comme *liste d'entiers*, il me faudrait écrire quelque chose comme `list(int)` ou encore `list[int]`, et cela serait source de confusion car ceci a déjà une signification dans le langage;
- de manière plus profonde, il faut distinguer entre `list` qui est un type concret (un objet qui sert à construire des instances), de `List` qui dans ce contexte doit plus être vu comme un type abstrait.

Pour bien voir cela, considérez l'exemple suivant :

```
from typing import Iterable
```

```
def lower_split(sep: str, inputs : Iterable[str]) -> str:
    return sep.join([x.lower() for x in inputs])
```

```
lower_split('--', ('AB', 'CD', 'EF'))
```

```
'ab--cd--ef'
```

On voit bien dans cet exemple que `Iterable` ne correspond pas à un type concret particulier, c'est un type abstrait dans le sens du *duck typing*.

un exemple plus complet

Voici un exemple tiré de la documentation du module `typing` qui illustre davantage de types construits à partir des types *builtin* du langage :

```
from typing import Dict, Tuple, List

ConnectionOptions = Dict[str, str]
Address = Tuple[str, int]
Server = Tuple[Address, ConnectionOptions]

def broadcast_message(message: str, servers: List[Server]) -> None:
    ...
```

```
# The static type checker will treat the previous type signature as
# being exactly equivalent to this one.
def broadcast_message(
    message: str,
    servers: List[Tuple[Tuple[str, int], Dict[str, str]]]) -> None:
    ...
```

J'en profite d'ailleurs (ça n'a rien à voir, mais...) pour vous signaler un objet python assez étrange :

```
# L'objet ... existe bel et bien en python
e1 = ...
e1
```

Ellipsis

qui sert principalement pour le slicing multi-dimensionnel de numpy. Mais ne nous égarons pas...

typage partiel

Puisque c'est un mécanisme optionnel, vous pouvez tout à fait ne typer qu'une partie de vos variables et paramètres :

```
# imaginez que vous ne typez pas n2, ni la valeur de retour
# c'est équivalent de dire ceci
def partially_typed(n1: int, n2):
    return None
```

```
# ou cela
from typing import Any

def partially_typed(n1: int, n2: Any) -> Any:
    return None
```

aliases

On peut facilement se définir des alias ; lorsque vous avez implémenté un système d'identifiants basé sur le type `int`, il est préférable de faire :

```
from typing import NewType

UserId = NewType('UserId', int)

user1_id : UserId = 0
```

plutôt que ceci, qui est beaucoup moins parlant :

```
user1_id : int = 0
```

70.2 Complément - niveau avancé

Pour ceux qui connaissent déjà la notion de classe (les autres peuvent ignorer la fin de ce complément) :

Grâce aux constructions `TypeVar` et `Generic`, il est possible de manipuler une notion de *variable de type*, que je vous montre sur un exemple tiré à nouveau de la documentation du module `typing` :


```
from typing import TypedDict, Generic
from logging import Logger

T = TypedDict('T')

class LoggedDict(Generic[T]):
    def __init__(self, value: T, name: str, logger: Logger) -> None:
        self.name = name
        self.logger = logger
        self.value = value

    def set(self, new: T) -> None:
        self.log('Set ' + repr(self.value))
        self.value = new

    def get(self) -> T:
        self.log('Get ' + repr(self.value))
        return self.value

    def log(self, message: str) -> None:
        self.logger.info('%s: %s', self.name, message)
```

qui vous donne je l'espère une idée de ce qu'il est possible de faire, et jusqu'où on peut aller avec les *type hints*. Si vous êtes intéressé par cette feature je vous invite à [poursuivre la lecture ici](#).

70.2.1 Pour en savoir plus

- la documentation officielle sur [le module typing](#) ;
- la page d'accueil [de l'outil mypy](#).
- le [PEP-525](#) sur le typage des paramètres et retours de fonctions, implémenté dans python-3.5 ;
- le [PEP-526](#) sur le typage des variables, implémenté dans 3.6.

Conditions & Expressions Booléennes

71.1 Complément : niveau basique

Nous présentons rapidement dans ce notebook comment construire la condition qui contrôle l'exécution d'un `if`

71.1.1 Tests considérés comme vrai

Lorsqu'on écrit une instruction comme

```
if <expression>:
    <do_something>
```

le résultat de l'expression peut **ne pas être un booléen**.

Par exemple, pour n'importe quel type numérique, la valeur 0 est considérée comme fausse. Cela signifie que

```
# ici la condition s'évalue à 0, donc on ne fait rien
if 3 - 3:
    print("ne passera pas par là")
```

```
# par contre si vous vous souvenez notre cours sur les flottants
# ici la condition donne un tout petit réel mais pas 0.
if 0.1 + 0.2 - 0.3:
    print("par contre on passe ici")
```

```
par contre on passe ici
```

De même, une chaîne vide, une liste vide, un tuple vide, sont considérés comme faux. Bref, vous voyez l'idée générale.

```
if "":
    print("ne passera pas par là")
if []:
    print("ne passera pas par là")
if ():
    print("ne passera pas par là")
```

```
# assez logiquement, None aussi
# est considéré comme faux
if None:
    print("ne passe toujours pas par ici")
```

71.1.2 Égalité

Les tests les plus simples se font à l'aide des opérateurs d'égalité, qui fonctionnent sur presque tous les objets. L'opérateur `==` vérifie si deux objets ont la même valeur :

```
bas = 12
haut = 25.82

# égalité ?
if bas == haut:
    print('==')
```

```
# non égalité ?
if bas != haut:
    print('!=')
```

```
!=
```

En général, deux objets de types différents ne peuvent pas être égaux.

```
# ces deux objets se ressemblent
# mais ils ne sont pas du même type !
if [1, 2] != (1, 2):
    print('!=')
```

```
!=
```

Par contre, des float, des int et des complex peuvent être égaux entre eux :

```
bas_reel = 12.
```

```
print(bas, bas_reel)
```

```
12 12.0
```

```
# le réel 12 et
# l'entier 12 sont égaux
if bas == bas_reel:
    print('int == float')
```

```
int == float
```

```
# ditto pour int et complex
if (12 + 0j) == 12:
    print('int == complex')
```

```
int == complex
```

Signalons à titre un peu anecdotique une syntaxe ancienne : historiquement et **seulement en python2** on pouvait aussi noter `<>` le test de non égalité. On trouve ceci dans du code ancien mais il faut éviter de l'utiliser :

```
%%python2

# l'ancienne forme de !=
if 12 <> 25:
    print("<> est obsolete et ne fonctionne qu'en python2")
```

```
<> est obsolete et ne fonctionne qu'en python2
```

71.1.3 Les opérateurs de comparaison

Sans grande surprise on peut aussi écrire

```
if bas <= haut:
    print('<=')
if bas < haut:
    print('<')
```

```
<=
<
```

```
if haut >= bas:
    print('>=')
if haut > bas:
    print('>')
```

```
>=
>
```

À titre de curiosité, on peut même écrire en un seul test une appartenance à un intervalle, ce qui donne un code plus lisible

```
x = (bas + haut) / 2
print(middle)
```

```
# deux tests en une expression
if bas <= x <= haut:
    print("dans l'intervalle")
```

```
dans l'intervalle
```

On peut utiliser les comparaisons sur une palette assez large de types, comme par exemple avec les listes

```
# on peut comparer deux listes, mais ATTENTION
[1, 2] <= [2, 3]
```

```
True
```

Il est parfois utile de vérifier le sens qui est donné à ces opérateurs selon le type; ainsi par exemple sur les ensembles - que nous verrons bientôt - ils se réfèrent à l'**inclusion** sur les ensembles.

Il faut aussi se méfier avec les types numériques, si un complexe est impliqué, comme dans l'exemple suivant. J'en profite pour anticiper un peu sur les exceptions, que nous verrons également plus tard :

```
# on ne peut pas par contre comparer deux nombres complexes

# l'instruction 'try .. except' permet d'attraper
# les erreurs à l'exécution
try:
```

```
2j <= 3j
except Exception as e:
    print("OOPS", type(e), e)
```

```
OOPS <class 'TypeError'> '<=' not supported between instances of 'complex' and
↳ 'complex'
```

71.1.4 Connecteurs logiques et / ou / non

On peut bien sûr combiner facilement plusieurs expressions entre elles, grâce aux opérateurs and, or et not

```
# il ne faut pas faire ceci, mettez des parenthèses
if 12 <= 25. or [1, 2] <= [2, 3] and not 12 <= 32 :
    print("OK mais pourrait être mieux")
```

```
OK mais pourrait être mieux
```

En termes de priorités : le plus simple si vous avez une expression compliquée reste de mettre les parenthèses qui rendent son évaluation claire et lisible pour tous. Aussi on préférera de beaucoup la formulation équivalente :

```
# c'est mieux
if 12 <= 25. or ([1, 2] <= [2, 3] and not 12 <= 32) :
    print("OK, c'est équivalent et plus clair")
```

```
OK, c'est équivalent et plus clair
```

```
# c'est bien le parenthésage ci-dessus, puisque:
if (12 <= 25. or [1, 2] <= [2, 3]) and not 12 <= 32 :
    print("ce n'est pas équivalent, ne passera pas par là")
```

71.1.5 Pour en savoir plus

Reportez-vous à la section sur les opérateurs booléens dans la documentation python.

72.1 Complément - niveau basique

72.1.1 Quels tests sont évalués ?

On a vu dans la vidéo que l’instruction conditionnelle `if` permet d’implémenter simplement des branchements à plusieurs choix, comme dans cet exemple :

```
s = 'berlin'
if 'a' in s:
    print('avec a')
elif 'b' in s:
    print('avec b')
elif 'c' in s:
    print('avec c')
else:
    print('sans a ni b ni c')
```

avec b

Comme on s’en doute, les expressions conditionnelles **sont évaluées jusqu’à obtenir un résultat vrai** - ou considéré comme vrai -, et le bloc correspondant est alors exécuté. Le point important ici est qu’**une fois qu’on a obtenu un résultat vrai**, on sort de l’expression conditionnelle **sans évaluer les autres conditions**. En termes savant, on parle d’évaluation paresseuse : on s’arrête dès qu’on peut.

Dans notre exemple, on aura évalué à la sortie `'a' in s`, et aussi `'b' in s`, mais pas `'c' in s`.

72.1.2 Pourquoi c’est important ?

C’est important de bien comprendre quels sont les tests qui sont réellement évalués pour deux raisons :

- d’abord, pour des raisons de performance ; comme on n’évalue que les tests nécessaires, si un des tests prend du temps, il est peut-être préférable de le faire en dernier ;
- mais aussi et surtout, il se peut tout à fait qu’un test fasse des **effets de bord**, c’est-à-dire qu’il modifie un ou plusieurs objets.

Dans notre premier exemple, les conditions elles-mêmes sont inoffensives ; la valeur de `s` reste *identique*, que l'on *évalue ou non* les différentes conditions.

Mais nous allons voir ci-dessous qu'il est relativement facile d'écrire des conditions qui **modifient** par **effet de bord** les objets mutables sur lesquelles elles opèrent, et dans ce cas il est crucial de bien assimiler la règle des évaluations des expressions dans un `if`.

72.2 Complément - niveau intermédiaire

72.2.1 Rappel sur la méthode `pop`

Pour illustrer la notion d'**effet de bord**, nous revenons sur la méthode de liste `pop()` qui, on le rappelle, renvoie un élément de liste **après l'avoir effacé** de la liste.

```
# on se donne une liste
liste = ['premier', 'deuxieme', 'troisieme']
print(f"liste={liste}")
```

```
liste=['premier', 'deuxieme', 'troisieme']
```

```
# pop(0) renvoie le premier element de la liste, et raccourcit la liste
element = liste.pop(0)
print(f"après pop(0), element={element} et liste={liste}")
```

```
après pop(0), element=premier et liste=['deuxieme', 'troisieme']
```

```
# et ainsi de suite
element = liste.pop(0)
print(f"après pop(0), element={element} et liste={liste}")
```

```
après pop(0), element=deuxieme et liste=['troisieme']
```

72.2.2 Conditions avec effet de bord

Une fois ce rappel fait, voyons maintenant l'exemple suivant :

```
liste = list(range(5))
print('liste en entree:', liste, 'de taille', len(liste))
```

```
liste en entree: [0, 1, 2, 3, 4] de taille 5
```

```
if liste.pop(0) <= 0:
    print('cas 1')
elif liste.pop(0) <= 1:
    print('cas 2')
elif liste.pop(0) <= 2:
    print('cas 3')
else:
    print('cas 4')
print('liste en sortie de taille', len(liste))
```

```
cas 1
liste en sortie de taille 4
```


Avec cette entrée, le premier test est vrai (car `pop(0)` renvoie 0), aussi on n'exécute en tout `pop()` qu'une seule fois, et donc à la sortie la liste n'a été raccourcie que d'un élément.

Exécutons à présent le même code avec une entrée différente :

```
liste = list(range(5, 10))
print('en entree: liste=', liste, 'de taille', len(liste))
```

```
en entree: liste= [5, 6, 7, 8, 9] de taille 5
```

```
if liste.pop(0) <= 0:
    print('cas 1')
elif liste.pop(0) <= 1:
    print('cas 2')
elif liste.pop(0) <= 2:
    print('cas 3')
else:
    print('cas 4')
print('en sortie: liste=', liste, 'de taille', len(liste))
```

```
cas 4
en sortie: liste= [8, 9] de taille 2
```

On observe que cette fois la liste a été **raccourcie 3 fois**, car les trois tests se sont révélés faux.

Cet exemple vous montre qu'il faut être attentif avec des conditions qui font des effets de bord. Bien entendu, ce type de pratique est de manière générale à utiliser avec beaucoup de discernement.

72.2.3 Court-circuit (*short-circuit*)

La logique que l'on vient de voir est celle qui s'applique aux différentes branches d'un `if` ; c'est la même logique qui est à l'œuvre aussi lorsque python évalue une condition logique à base de `and` et `or`. C'est ici aussi une forme d'évaluation paresseuse.

Pour illustrer cela, nous allons nous définir deux fonctions toutes simples qui renvoient `True` et `False` mais avec une impression de sorte qu'on voit lorsqu'elles sont exécutées :

```
def true():
    print('true')
    return True
```

```
def false():
    print('false')
    return False
```

```
true()
```

```
true
```

```
True
```

Ceci va nous permettre d'illustrer notre point, qui est que lorsque python évalue un `and` ou un `or`, il **n'évalue la deuxième condition que si c'est nécessaire**. Ainsi par exemple :

```
false() and true()
```

```
false
```

False

Dans ce cas, python évalue la première partie du `and` - qui provoque l'impression de `false` - et comme le résultat est faux, il n'est **pas nécessaire** d'évaluer la seconde condition, on sait que de toute façon le résultat du `and` est forcément faux. C'est pourquoi vous ne voyez pas l'impression de `true`.

De manière symétrique avec un `or` :

```
true() or false()
```

```
true
```

True

À nouveau ici il n'est pas nécessaire d'évaluer `false()`, et donc seul `true` est imprimé à l'évaluation.

À titre d'exercice, essayez de dire combien d'impressions sont émises lorsqu'on évalue cette expression un peu plus compliquée :

```
true() and (false() or true()) or (true () and false())
```

```
true  
false  
true
```

True

Une forme alternative du `if`

73.1 Complément - niveau basique

73.1.1 Expressions et instructions

Les constructions python que nous avons vues jusqu'ici peuvent se ranger en deux familles :

- d'une part les **expressions** sont les fragments de code qui **retournent une valeur** ;
 - c'est le cas lorsqu'on invoque n'importe quel opérateur numérique, pour les appels de fonction, ...
- d'autre part les **instructions** ;
 - dans cette famille, jusqu'ici nous avons vu l'affectation, et `if`, et nous en verrons bien d'autres.

La différence essentielle est que les expressions peuvent être combinées entre elles pour faire des expressions arbitrairement grosses. Aussi, si vous avez un doute pour savoir si vous avez affaire à une expression ou à une instruction, demandez vous si vous pourriez utiliser ce code **comme membre droit d'une affectation**. Si oui, vous avez une expression.

73.1.2 `if` est une instruction

La forme du `if` qui vous a été présentée pendant la vidéo ne peut pas servir à renvoyer une valeur, c'est donc une **instruction**.

Imaginons maintenant qu'on veuille écrire quelque chose d'aussi simple que “affecter à `y` la valeur 12 ou 35, selon que `x` est vrai ou non”.

Avec les notions introduites jusqu'ici, il nous faudrait écrire ceci :

```
x = True # ou quoi que ce soit d'autre
if x:
    y = 12
else:
    y = 35
print(y)
```

```
12
```

73.1.3 Expression conditionnelle

Il existe aussi en python une expression qui fait le même genre de test; c'est la forme dite d'**expression conditionnelle**, qui est une **expression à part entière**, avec la syntaxe :

```
<resultat_si_vrai> if <condition> else <resultat_si_faux>
```

Ainsi on pourrait écrire l'exemple ci-dessus de manière plus simple et plus concise comme ceci

```
y = 12 if x else 35
print(y)
```

```
12
```

Cette construction peut souvent rendre le style de programmation plus fonctionnel et plus fluide.

73.2 Complément - niveau intermédiaire

73.2.1 Imbrications

Puisque cette forme est une expression, on peut l'utiliser .. dans une autre expression conditionnelle, comme ici :

```
# on veut calculer en fonction d'une entrée x
# une sortie qui vaudra
# -1 si x < -10
# 0 si -10 <= x <= 10
# 1 si x > 10

x = 5 # ou quoi que ce soit d'autre

valeur = -1 if x < -10 else 0 if x <= 10 else 1

print(valeur)
```

```
0
```

Remarquez bien que dans cet exemple la ligne

```
valeur = -1 if x < -10 else 0 if x <= 10 else 1
```

se lit en réalité

```
valeur = -1 if x < -10 else (0 if x <= 10 else 1)
```

Pour en savoir plus ~~~~~

- La section sur les [expressions conditionnelles](#) de la documentation python
- Le [PEP308](#) qui résume les discussions ayant donné lieu au choix de la syntaxe adoptée

De manière générale, les PEP rassemblent les discussions préalables à toutes les évolutions majeures du langage python.

Récapitulatif sur les conditions dans un `if`

74.1 Complément - niveau basique

Dans ce complément nous résumons ce qu'il faut savoir pour écrire une condition dans un `if`.

74.1.1 Expression vs instruction

Nous avons déjà introduit la différence entre instruction et expression, lorsque nous avons vu l'expression conditionnelle :

- une expression est un fragment de code qui “retourne quelque chose”,
- alors qu'une instruction permet bien souvent de faire une action, mais ne retourne rien.

Ainsi parmi les notions que nous avons vues jusqu'ici, nous pouvons citer dans un ordre arbitraire :

| Instructions | Expressions |
|------------------------------|--|
| affectation | appel de fonction |
| <code>import</code> | opérateurs <code>is</code> , <code>in</code> , <code>==</code> , ... |
| instruction <code>if</code> | expression conditionnelle |
| instruction <code>for</code> | compréhension(s) |

74.1.2 Toutes les expressions sont éligibles

Comme condition d'une instruction `if`, on peut mettre n'importe quelle expression. On l'a déjà signalé, il n'est pas nécessaire que cette expression retourne un booléen :

```
# dans ce code le test
# if n % 3:
# est équivalent à
# if n % 3 != 0:

for n in (18, 19):
    if n % 3:
        print(f"{n} non divisible par trois")
```

```
else:
    print(f"{n} divisible par trois")
```

```
18 divisible par trois
19 non divisible par trois
```

74.1.3 Une valeur est-elle “vraie” ?

Se pose dès lors la question de savoir précisément quelles valeurs sont considérées comme *vraies* par l’instruction `if`.

Parmi les types de base, nous avons déjà eu l’occasion de l’évoquer, les valeurs *fausses* sont typiquement

- 0 pour les valeurs numériques
- les objets vides pour les chaînes, listes, ensembles, dictionnaires, etc.

Pour en avoir le cœur net, pensez à utiliser dans le terminal interactif la fonction `bool`. Comme pour toutes les fonctions qui portent le nom d’un type, la fonction `bool` est un constructeur qui fabrique un objet booléen.

Si vous appelez `bool` sur un objet, la valeur de retour - qui est donc par construction une valeur booléenne - vous indique, cette fois sans ambiguïté - comment se comportera `if` avec cette entrée.

```
def show_bool(x):
    print(f"condition {repr(x):>10} considérée comme {bool(x)}")
```

```
for exp in [None, "", 'a', [], [1], (), (1, 2), {}, {'a': 1}, set(), {1}]:
    show_bool(exp)
```

```
condition      None considérée comme False
condition      '' considérée comme False
condition      'a' considérée comme True
condition      [] considérée comme False
condition      [1] considérée comme True
condition      () considérée comme False
condition      (1, 2) considérée comme True
condition      {} considérée comme False
condition      {'a': 1} considérée comme True
condition      set() considérée comme False
condition      {1} considérée comme True
```

74.1.4 Quelques exemples d’expressions

Référence à une variable et dérivés

```
a = list(range(4))
print(a)
```

```
[0, 1, 2, 3]
```

```
if a:
    print("a n'est pas vide")
if a[0]:
    print("on ne passe pas par ici")
if a[1]:
    print("a[1] n'est pas nul")
```

```
a n'est pas vide
a[1] n'est pas nul
```

Appels de fonction ou de méthode

```
chaine = "jean"
if chaine.upper():
    print("la chaine mise en majuscule n'est pas vide")
```

```
la chaine mise en majuscule n'est pas vide
```

```
# on rappelle qu'une fonction qui ne fait pas 'return' retourne None
def procedure(a, b, c):
    "cette fonction ne retourne rien"
    pass

if procedure(1, 2, 3):
    print("ne passe pas ici car procedure retourne None")
else:
    print("par contre on passe ici")
```

```
par contre on passe ici
```

Compréhensions

Il découle de ce qui précède qu'on peut tout à fait mettre une compréhension comme condition, ce qui peut être utile pour savoir si au moins un élément remplit une condition, comme par exemple :

```
inputs = [23, 65, 24]

# y a-t-il dans inputs au moins un nombre
# dont le carré est de la forme 10*n+5
def condition(n):
    return (n * n) % 10 == 5

if [input for input in inputs if condition(input)]:
    print("au moins une entrée convient")
```

```
au moins une entrée convient
```

Opérateurs

Nous avons déjà eu l'occasion de rencontrer la plupart des opérateurs de comparaison du langage, dont voici à nouveau les principaux

| Famille | Exemples |
|--------------|--------------------|
| Égalité | ==, !=, is, is not |
| Appartenance | in |
| Comparaison | <=, <, >, >= |
| Logiques | and, or, not |

74.2 Complément - niveau intermédiaire

74.2.1 Remarques sur les opérateurs

Voici enfin quelques remarques sur ces opérateurs

opérateur d'égalité ==

L'opérateur == ne fonctionne en général (sauf pour les nombres) que sur des objets de même type ; c'est-à-dire que notamment un tuple ne sera jamais égal à une liste :

```
[ ] == ( )
```

```
False
```

```
[1, 2] == (1, 2)
```

```
False
```

opérateur logiques

Comme c'est le cas avec par exemple les opérateurs arithmétiques, les opérateurs logiques ont une *priorité*, qui précise le sens des phrases non parenthésées. C'est-à-dire pour être explicite, que de la même manière que

```
12 + 4 * 8
```

est équivalent à

```
12 + ( 4 * 8 )
```

pour les booléens il existe une règle de ce genre et

```
a and not b or c and d
```

est équivalent à

```
(a and (not b)) or (c and d)
```

Mais en fait, il est assez facile de s'emmêler dans ces priorités, et c'est pourquoi il est **très fortement conseillé** de parenthéser.

opérateurs logiques (2)

Remarquez aussi que les opérateurs logiques peuvent être appliqués à des valeurs qui ne sont pas booléennes :

```
2 and [1, 2]
```

```
[1, 2]
```

```
None or "abcde"
```

```
'abcde'
```


Dans la logique de l'évaluation paresseuse qu'on a vue récemment, remarquez que lorsque l'évaluation d'un `and` ou d'un `or` ne peut pas être court-circuitée, le résultat est alors toujours le résultat de la dernière expression évaluée :

```
1 and 2 and 3
```

```
3
```

```
1 and 2 and 3 and '' and 4
```

```
''
```

```
[] or "" or {}
```

```
{}
```

```
[] or "" or {} or 4 or set()
```

```
4
```

74.2.2 Expression conditionnelle dans une instruction `if`

En toute rigueur on peut aussi mettre un `<> if <> else <>` - donc une expression conditionnelle - comme condition dans une instruction `if`. Nous le signalons pour bien illustrer la logique du langage, mais cette pratique n'est bien sûr pas du tout conseillée.

```
# cet exemple est volontairement tiré par les cheveux
# pour bien montrer qu'on peut mettre n'importe quelle expression comme condition
a = 1
# ceci est franchement illisible
if 0 if not a else 2:
    print("une construction illisible")
# et encore pire
if 0 if a else 3 if a + 1 else 2:
    print("encore pire")
```

```
une construction illisible
```

74.2.3 Pour en savoir plus

<https://docs.python.org/3/tutorial/datastructures.html#more-on-conditions>

74.2.4 Types définis par l'utilisateur

Pour anticiper un tout petit peu, nous verrons que les classes en python vous donnent le moyen de définir vos propres types d'objets. Nous verrons à cette occasion qu'il est possible d'indiquer à python quels sont les objets de type `MaClasse` qui doivent être considérés comme `True` ou comme `False`.

De manière plus générale, tous les traits natifs du langage sont redéfinissables sur les classes. Nous verrons par exemple également comment donner du sens à des phrases comme

```
mon_objet = MaClasse()
if mon_objet:
    <faire quelque chose>
```

ou encore

```
mon_objet = MaClasse()  
for partie in mon_objet:  
    <faire quelque chose sur partie>
```

Mais n'anticipons pas trop, rendez-vous en semaine 6.

L'instruction `if`

75.1 Exercice - niveau basique

75.1.1 Répartiteur (1)

```
# on charge l'exercice
from exo_dispatch import exo_dispatch1
```

On vous demande d'écrire une fonction `dispatch1`, qui prend en argument deux entiers `a` et `b`, et qui renvoie selon les cas :

| | <i>a pair</i> | <i>a impair</i> |
|-----------------|---------------|-----------------|
| <i>b pair</i> | $a^2 + b^2$ | $(a - 1) * b$ |
| <i>b impair</i> | $a * (b - 1)$ | $a^2 - b^2$ |

```
# un petit exemples
exo_dispatch1.example()
```

| Appel | Résultat Attendu |
|------------------------------|------------------|
| <code>dispatch1(3, 7)</code> | <code>-40</code> |

```
def dispatch1(a, b):
    "<votre_code>"
```

```
# pour vérifier votre code
exo_dispatch1.correction(dispatch1)
```

75.2 Exercice - niveau basique

75.2.1 Répartiteur (2)

```
# chargement de l'exercice
from exo_dispatch import exo_dispatch2
```

Dans une seconde version de cet exercice, on vous demande d'écrire une fonction `dispatch2` qui prend en arguments

- a et b deux entiers
- **A et B deux ensembles (chacun pouvant être** matérialisé par un ensemble, une liste ou un tuple) et qui renvoie selon les cas

| | $a \in A$ | $a \notin A$ |
|--------------|---------------|---------------|
| $b \in B$ | $a^2 + b^2$ | $(a - 1) * b$ |
| $b \notin B$ | $a * (b - 1)$ | $a^2 + b^2$ |

```
def dispatch2(a, b, A, B):
    "<votre_code>"
```

```
# pour vérifier votre code
exo_dispatch2.correction(dispatch2)
```

Expression conditionnelle

76.1 Exercice - niveau basique

76.1.1 Analyse et mise en forme

```
# Pour charger l'exercice
from corrections.exo_libelle import exo_libelle
```

Un fichier contient, dans chaque ligne, des informations (champs) séparées par des virgules. Les espaces et tabulations présents dans la ligne ne sont pas significatifs et doivent être ignorés.

Dans cet exercice de niveau basique, on suppose que chaque ligne a exactement 3 champs, qui représentent respectivement le nom, le prénom, et le rang d'une personne dans un classement. Une fois les espaces et tabulations ignorés, on ne fait pas de vérification sur le contenu des 3 champs.

On vous demande d'écrire la fonction `libelle`, qui sera appelée pour chaque ligne du fichier. Cette fonction :

- prend en argument une ligne (chaîne de caractères)
- retourne une chaîne de caractères mise en forme (voir plus bas)
- ou bien retourne `None` si la ligne n'a pas pu être analysée, parce qu'elle ne vérifie pas les hypothèses ci-dessus (c'est notamment le cas si on ne trouve pas exactement les 3 champs)

La mise en forme consiste à retourner

```
Nom.Prenom (message)
```

le *message* étant lui-même le *rang* mis en forme pour afficher '1er', '2nd' ou '*n*-ème' selon le cas. Voici quelques exemples

```
# voici quelques exemples de ce qui est attendu
exo_libelle.example()
```

| Arguments Résultat Attendu | |
|---|--------------------------|
| “Joseph, Dupont, 4” “Dupont.Joseph (4-ème)” | |
| “Jean” | None |
| “Jules , Durand, 1” | “Durand.Jules (1er)” |
| “ Ted, Mosby, 2,” | None |
| “ Jacques , Martin, 3 t” | “Martin.Jacques (3-ème)” |
| “Sheldon, Cooper ,5, “ | None |
| “t John, Doet, “ | “Doe.John (-ème)” |
| “John, Smith, , , 3” | None |

```
# écrivez votre code ici
def libelle(ligne):
    "<votre_code>"
```

```
# pour le vérifier
exo_libelle.correction(libelle)
```

La boucle `while .. else`

77.1 Complément - niveau basique

77.1.1 Boucles sans fin - `break`

Utiliser `while` plutôt que `for` est une affaire de style et d'habitude. Cela dit en python, avec les notions d'itérable et d'itérateur, on a tendance à privilégier l'usage du `for` pour les boucles finies et déterministes.

Le `while` reste malgré tout d'un usage courant, et notamment avec une condition `True`.

Par exemple le code de l'interpréteur interactif de python pourrait ressembler, vu de très loin, à quelque chose comme ceci

```
while True:
    print(eval(read()))
```

Notez bien par ailleurs que les instructions `break` et `continue` fonctionnent, à l'intérieur d'une boucle `while`, exactement comme dans un `for`, c'est-à-dire que :

- `continue` termine l'itération courante mais reste dans la boucle, alors que
- `break` interrompt l'itération courante et sort également de la boucle.

77.2 Complément - niveau intermédiaire

77.2.1 Rappel sur les conditions

On peut utiliser dans une boucle `while` toutes les formes de conditions que l'on a vues à l'occasion de l'instruction `if`.

Dans le contexte de la boucle `while` on comprend mieux, toutefois, pourquoi le langage autorise d'écrire des conditions dont le résultat n'est **pas nécessairement un booléen**. Voyons cela sur un exemple simple :

```
# une autre façon de parcourir une liste
liste = ['a', 'b', 'c']

while liste:
    element = liste.pop()
    print(element)
```

```
c
b
a
```

77.2.2 Une curiosité : la clause `else`

Signalons enfin que la boucle `while` - au même titre d'ailleurs que la boucle `for`, peut être assortie d'une clause `'else'` <https://docs.python.org/3/reference/compound_stmts.html#the-while-statement>, qui est exécutée à la fin de la boucle, **sauf dans le cas d'une sortie avec `"break"`**

```
# Un exemple de while avec une clause else

# si break_mode est vrai on va faire un break
# après le premier élément de la liste
def scan(liste, break_mode):

    # un message qui soit un peu parlant
    message = "avec break" if break_mode else "sans break"
    print(message)
    while liste:
        print(liste.pop())
        if break_mode:
            break
    else:
        print('else...')
```

```
# sortie de la boucle sans break:
# on passe par else
scan(['a'], False)
```

```
sans break
a
else...
```

```
# on sort de la boucle par le break
scan(['a'], True)
```

```
avec break
a
```

Ce trait est toutefois **très rarement** utilisé.

Calculer le PGCD

78.1 Exercice - niveau basique

```
# chargement de l'exercice
from corrections.exo_pgcd import exo_pgcd
```

On vous demande d'écrire une fonction qui calcule le pgcd de deux entiers, en utilisant l'algorithme d'Euclide.

Les deux paramètres sont supposés être des entiers positifs ou nuls (pas la peine de le vérifier). Dans le cas où un des deux paramètres est nul, on décide par convention de retourner la valeur 0.

Remarque on peut tout à fait utiliser une fonction récursive pour implémenter l'algorithme d'Euclide. Par exemple cette version de pgcd fonctionne très bien aussi (en supposant $a \geq b$)

```
def pgcd(a, b):
    "Le pgcd avec une fonction récursive"
    if not b:
        return a
    return pgcd(b, a % b)
```

Cependant, il vous est demandé ici d'utiliser une boucle `while`, qui est le sujet de la séquence, pour implémenter pgcd.

```
# à vous de jouer
def pgcd(a, b):
    "<votre code>"
```

```
# pour vérifier votre code
exo_pgcd.correction(pgcd)
```


Exercice

79.1 Niveau basique

```
from corrections.exo_taxes import exo_taxes
```

On se propose d'écrire une fonction `taxes` qui calcule le montant de l'impôt sur le revenu au Royaume-uni.

Le barème est [publié ici par le gouvernement anglais](#), voici les données utilisées pour l'exercice :

| Tranche | Revenu imposable | Taux |
|---------------------|---------------------|------|
| Non imposable | jusque £11.500 | 0% |
| Taux de base | £11.501 à £45.000 | 20% |
| Taux élevé | £45.001 à £150.000 | 40% |
| Taux supplémentaire | au delà de £150.000 | 45% |

Donc naturellement il s'agit d'écrire une fonction qui prend en argument le revenu imposable, et retourne le montant de l'impôt, **arrondi à l'entier inférieur**.

```
exo_taxes.example()
```

| Appel Résultat Attendu | |
|---------------------------|------|
| <code>taxes(0)</code> | 0 |
| <code>taxes(45000)</code> | 6700 |

Indices

- évidemment on parle ici d'une fonction continue ;
- aussi en termes de programmation, je vous encourage à séparer la définition des tranches de la fonction en elle-même.

```
def taxes(income):
    # ce n'est pas la bonne réponse
    return (income-11_500) * (20/100)
```

```
exo_taxes.correction(taxes)
```

79.1.1 Représentation graphique

Comme d'habitude vous pouvez voir la représentation graphique de votre fonction :

```
import numpy as np
import matplotlib.pyplot as plt
```

```
%matplotlib inline
plt.ion()
```

```
X = np.linspace(0, 200_000)
Y = [taxes(x) for x in X]
plt.plot(X, Y);
```

```
# et pour changer la taille de la figure
%pylab inline
pylab.rcParams['figure.figsize'] = (10, 8)
plt.plot(X, Y);
```

Le scope `builtin`

80.1 Complément - niveau avancé

80.1.1 Ces noms qui viennent de nulle part

Nous avons vu déjà un certain nombre de **fonctions** **built-in** comme par exemple

```
open, len, zip
```

```
(<function io.open>, <function len>, zip)
```

Ces noms font partie du **module** “**builtins**”. Il est cependant particulier puisque tout se passe **comme si** on avait fait avant toute chose :

```
from builtins import *
```

sauf que cet import est implicite.

80.1.2 On peut réaffecter un nom *builtin*

Quoique ce soit une pratique déconseillée, il est tout à fait possible de redéfinir ces noms ; on peut faire par exemple

```
def open(encoding='utf-8', *args):  
    print("ma fonction open")  
    pass
```

qui est naturellement **très vivement déconseillé**.

80.1.3 On ne peut pas réaffecter un mot clé

À titre de digression, rappelons que les noms prédéfinis dans le module `builtins` sont, à cet égard aussi, très différents des mots-clés comme `if`, `def`, `with` et autres `for` qui eux, ne peuvent pas être modifiés en aucune manière :

```
>>> lambda = 1
      File "<stdin>", line 1
        lambda = 1
            ^
SyntaxError: invalid syntax
```

80.1.4 Retrouver un objet *builtin*

Même quand on sait qu'il faut éviter de redéfinir un nom prédéfini dans `builtins`, on n'a pas toujours en tête la liste exhaustive des noms prédéfinis, aussi il arrive qu'on le fasse involontairement (même si un bon éditeur de texte vous signalera les fonctions *built-in* avec une coloration syntaxique spécifique).

Sachez que vous pouvez toujours "retrouver" alors la fonction prédéfinie en l'important explicitement du module `builtins`. Par exemple, pour réaliser notre ouverture de fichier, nous pouvons toujours faire :

```
# nous ne pouvons pas utiliser open puisque
open()
```

```
ma fonction open
```

```
# pour être sûr d'utiliser la bonne fonction open

import builtins

with builtins.open("builtin.txt", "w", encoding="utf-8") as f:
    f.write("quelque chose")
```

Ou encore, de manière équivalente :

```
from builtins import open as builtin_open

with builtin_open("builtin.txt", "r", encoding="utf-8") as f:
    print(f.read())
```

```
quelque chose
```

80.1.5 Liste des fonctions prédéfinies

Vous pouvez trouver la liste des fonctions prédéfinies avec la fonction `dir` sur le module `builtins` comme ci-dessous (qui vous montre aussi les exceptions prédéfinies, qui commencent par une majuscule), ou dans la documentation sur [les fonctions prédéfinies](#) :

```
dir(builtins)
```

Vous remarquez que les exceptions (les symboles qui commencent par des majuscules) représentent à elles-seules une proportion substantielle de cet espace de noms.

Visibilité des variables de boucle

81.1 Complément - niveau basique

81.1.1 Une astuce

Dans ce complément, nous allons beaucoup jouer avec le fait qu'une variable soit définie ou non. Pour nous simplifier la vie, et surtout rendre les cellules plus indépendantes les unes des autres si vous devez les rejouer, nous allons utiliser la formule un peu magique suivante :

```
# on détruit la variable i si elle existe
if 'i' in locals():
    del i
```

qui repose d'une part sur l'instruction `del` que nous avons déjà vue, et sur la fonction *builtin* `locals` que nous verrons plus tard ; cette formule a l'avantage qu'on peut l'exécuter dans n'importe quel contexte, que `i` soit définie ou non.

81.1.2 Une variable de boucle reste définie au-delà de la boucle

Une variable de boucle est définie (assignée) dans la boucle et **reste *visible*** une fois la boucle terminée. Le plus simple est de le voir sur un exemple :

```
# La variable 'i' n'est pas définie
try:
    i
except NameError as e:
    print('OOPS', e)
```

```
OOPS name 'i' is not defined
```

```
# si à présent on fait une boucle
# avec i comme variable de boucle
for i in [0]:
    pass
```

```
# alors maintenant i est définie
i
```

```
0
```

On dit que la variable *fuite* (en anglais “*leak*”), dans ce sens qu’elle continue d’exister au delà du bloc de la boucle à proprement parler.

On peut être tenté de tirer profit de ce trait, en lisant la valeur de la variable après la boucle ; l’objet de ce complément est de vous inciter à la prudence, et d’attirer votre attention sur certains points qui peuvent être sources d’erreur.

81.1.3 Attention aux boucles vides

Tout d’abord, il faut faire attention à ne pas écrire du code qui dépende de ce trait **si la boucle peut être vide**. En effet, si la boucle ne s’exécute pas du tout, la variable n’est **pas affectée** et donc elle n’est **pas définie**. C’est évident, mais ça peut l’être moins quand on lit du code réel, comme par exemple :

```
# on détruit la variable i si elle existe
if 'i' in locals():
    del i
```

```
# une façon très scabreuse de calculer la longueur de l
def length(l):
    for i, x in enumerate(l):
        pass
    return i + 1

length([1, 2, 3])
```

```
3
```

Ça a l’air correct, sauf que :

```
length([])
```

```
-----
UnboundLocalError                                Traceback (most recent call last)

<ipython-input-6-8c0f554916d9> in <module>()
----> 1 length([])

<ipython-input-5-54c4139d6f55> in length(l)
      3     for i, x in enumerate(l):
      4         pass
----> 5     return i + 1
      6
      7 length([1, 2, 3])

UnboundLocalError: local variable 'i' referenced before assignment
```

Ce résultat mérite une explication. Nous allons voir très bientôt l’exception `UnboundLocalError`, mais pour le moment sachez qu’elle se produit lorsqu’on a dans une fonction une variable locale et une variable globale de même nom. Alors, pourquoi l’appel `length([1, 2, 3])` retourne-t-il sans encombre, alors que pour l’appel `length([])` il y a une exception ? Cela est lié à la manière dont python détermine qu’une variable est locale.

Une variable est locale dans une fonction si elle est assignée dans la fonction explicitement (avec une opération d'affectation) ou implicitement (par exemple avec une boucle `for` comme ici); nous reviendrons sur ce point un peu plus tard. Mais pour les fonctions, pour une raison d'efficacité, une variable est définie comme locale à la phase de pré-compilation, c'est-à-dire *avant* l'exécution du code. Le pré-compilateur ne peut pas savoir quel sera l'argument passé à la fonction, il peut simplement savoir qu'il y a une boucle `for` utilisant la variable `i`, il en conclut que `i` est locale pour toute la fonction.

Lors du premier appel, on passe une liste à la fonction, liste qui est parcourue par la boucle `for`. En sortie de boucle, on a bien une variable locale `i` qui vaut 3. Lors du deuxième appel par contre, on passe une liste vide à la fonction, la boucle `for` ne peut rien parcourir, donc elle termine immédiatement. Lorsque l'on arrive à la ligne `return i + 1` de la fonction, la variable `i` n'a pas de valeur (on doit donc chercher `i` dans le module), mais `i` a été définie par le pré-compilateur comme étant locale, on a donc dans la même fonction une variable `i` locale et une référence à une variable `i` globale, ce qui provoque l'exception `UnboundLocalError`.

81.1.4 Comment faire alors ?

Utiliser une autre variable

La première voie consiste à déclarer une variable externe à la boucle et à l'affecter à l'intérieur de la boucle, c'est-à-dire :

```
# on veut chercher le premier de ces nombres dont le carré est un multiple de 5
candidates = [3, -15, 1, 8]

def checks(candidate):
    return candidate ** 2 % 5 == 0
```

```
# plutôt que de faire ceci
for item in candidates:
    if checks(item):
        break
print('trouvé solution', item)
```

```
trouvé solution -15
```

```
# il vaut mieux faire ceci
solution = None
for item in candidates:
    if checks(item):
        solution = item
        break

print('trouvé solution', solution)
```

```
trouvé solution -15
```

Au minimum initialiser la variable

Au minimum, si vous utilisez la variable de boucle après la boucle, il est vivement conseillé de l'**initialiser** explicitement **avant** la boucle, pour vous prémunir contre les boucles vides, comme ceci :

```
# la fonction length de tout à l'heure
def length1(l):
    for i, x in enumerate(l):
        pass
    return i + 1
```

```
# une version plus robuste
def length2(l):
    # on initialise i explicitement
    # pour le cas où l est vide
    i = -1
    for i, x in enumerate(l):
        pass
    # comme cela i est toujours déclarée
    return i + 1
```

```
length1([])
```

```
-----
UnboundLocalError                                Traceback (most recent call last)

<ipython-input-12-ebdd61d1db90> in <module>()
----> 1 length1([])

<ipython-input-10-6a59feac567b> in length1(l)
      3     for i, x in enumerate(l):
      4         pass
----> 5     return i + 1

UnboundLocalError: local variable 'i' referenced before assignment
```

```
length2([])
```

```
0
```

81.1.5 Les compréhensions

Notez bien que par contre, les variables de compréhension **ne fuient pas** (contrairement à ce qui se passait en python2) :

```
# on détruit la variable i si elle existe
if 'i' in locals():
    del i
```

```
# en python3, les variables de compréhension ne fuient pas
[i**2 for i in range(3)]
```

```
[0, 1, 4]
```

```
# ici i est à nouveau indéfinie
try:
    i
except NameError as e:
    print("OOPS", e)
```

```
OOPS name 'i' is not defined
```

L'exception `UnboundLocalError`

82.1 Complément - niveau intermédiaire

Nous résumons ici quelques cas simples de portée de variables.

82.1.1 Variable locale

Les **arguments** attendus par la fonction sont considérés comme des variables **locales**, c'est-à-dire dans l'espace de noms de la fonction.

Pour définir une autre variable locale, il suffit de la définir (l'affecter), elle devient alors accessible en lecture :

```
def ma_fonction1():  
    variable1 = "locale"  
    print(variable1)  
  
ma_fonction1()
```

```
locale
```

et ceci que l'on ait ou non une variable globale de même nom

```
variable2 = "globale"  
  
def ma_fonction2():  
    variable2 = "locale"  
    print(variable2)  
  
ma_fonction2()
```

```
locale
```

82.1.2 Variable globale

On peut accéder **en lecture** à une variable globale sans précaution particulière :

```
variable3 = "globale"

def ma_fonction3():
    print(variable3)

ma_fonction3()
```

```
globale
```

82.1.3 Mais il faut choisir !

Par contre on ne **peut pas** faire la chose suivante dans une fonction. On ne peut pas utiliser **d'abord** une variable comme une variable **globale**, puis essayer de l'affecter localement - ce qui signifie la déclarer comme une **locale** :

```
# cet exemple ne fonctionne pas et leve UnboundLocalError
variable4 = "globale"

def ma_fonction4():
    # on référence la variable globale
    print(variable4)
    # et maintenant on crée une variable locale
    variable4 = "locale"

# on "attrape" l'exception
try:
    ma_fonction4()
except Exception as e:
    print(f"OOPS, exception {type(e)}:\n{e}")
```

```
OOPS, exception <class 'UnboundLocalError':
local variable 'variable4' referenced before assignment
```

82.1.4 Comment faire alors ?

L'intérêt de cette erreur est d'interdire de mélanger des variables locales et globales de même nom dans une même fonction. On voit bien que ça serait vite incompréhensible. Donc une variable dans une fonction peut être **ou bien** locale si elle est affectée dans la fonction **ou bien** globale, mais **pas les deux à la fois**. Si vous avez une erreur `UnboundLocalError`, c'est qu'à un moment donné vous avez fait cette confusion.

Vous vous demandez peut-être à ce stade, mais comment fait-on alors pour modifier une variable globale depuis une fonction ? Pour cela il faut utiliser l'instruction `global` comme ceci :

```
# Pour résoudre ce conflit il faut explicitement
# déclarer la variable comme globale
variable5 = "globale"

def ma_fonction5():
    global variable5
    # on référence la variable globale
    print("dans la fonction", variable5)
    # cette fois on modifie la variable globale
    variable5 = "changée localement"

ma_fonction5()
print("après la fonction", variable5)
```

```
dans la fonction globale
après la fonction changée localement
```

Nous reviendrons plus longuement sur l'instruction `global` dans la prochaine vidéo.

82.1.5 Bonnes pratiques

Cela étant dit, l'utilisation de variables globales est généralement considérée comme une mauvaise pratique.

Le fait d'utiliser une variable globale en *lecture seule* peut rester acceptable, lorsqu'il s'agit de matérialiser une constante qu'il est facile de changer. Mais dans une application aboutie, ces constantes elles-mêmes peuvent être modifiées par l'utilisateur via un système de configuration, donc on préférera passer en argument un objet *config*.

Et dans les cas où votre code doit recourir à l'utilisation de l'instruction `global`, c'est très probablement que quelque chose peut être amélioré au niveau de la conception de votre code.

Il est recommandé, au contraire, de passer en argument à une fonction tout le contexte dont elle a besoin pour travailler ; et à l'inverse d'utiliser le résultat d'une fonction plutôt que de modifier une variable globale.

Les fonctions `globals` et `locals`

83.1 Complément - niveau intermédiaire

83.1.1 Un exemple

python fournit un accès à la liste des noms et valeurs des variables visibles à cet endroit du code. Dans le jargon des langages de programmation on appelle ceci l'**environnement**.

Cela est fait grâce aux fonctions *builtin* `globals` et `locals`, que nous allons commencer par essayer sur quelques exemples. Nous avons pour cela écrit un module dédié :

```
import env_locals_globals
```

Dont voici le code

```
from modtools import show_module
show_module(env_locals_globals)
```

Fichier E:python_code\python-master\w4env_locals_globals.py

```
-----
| # -- coding: utf-8 --
|
| """
| un module pour illustrer les fonctions globals et locals
| """
|
| globale = "variable globale au module"
|
| def display_env(env):
|     """
|     affiche un environnement
|     on affiche juste le nom et le type de chaque variable
|     """
|     for variable, valeur in sorted(env.items()):
|         print("{:>20} â†' {}".format(variable, type(valeur).__name__))
|
| def temoin(x):
```

```
|     "la fonction témoin"
|     y = x ** 2
|     print(20 * '-', 'globals:')
|     display_env(globals())
|     print(20 * '-', 'locals:')
|     display_env(locals())
|
| class Foo:
|     "une classe vide"
```

et voici ce qu'on obtient lorsqu'on appelle

```
env_locals_globals.témoin(10)
```

```
----- globals:
          Foo → type
    __builtins__ → dict
    __cached__ → str
    __doc__ → str
    __file__ → str
    __loader__ → SourceFileLoader
    __name__ → str
    __package__ → str
    __spec__ → ModuleSpec
    display_env → function
    globale → str
    témoin → function
----- locals:
          x → int
          y → int
```

83.1.2 Interprétation

Que nous montre cet exemple ?

- D'une part la fonction **"globals"** nous donne la liste des symboles définis au niveau de **l'espace de noms du module**. Il s'agit évidemment du module dans lequel est définie la fonction, pas celui dans lequel elle est appelée. Vous remarquerez que ceci englobe **tous** les symboles du module `env_locals_globals`, et non pas seulement ceux définis avant `témoin`, c'est-à-dire la variable `globale`, les deux fonctions `display_env` et `témoin`, et la classe `Foo`.
- D'autre part **"locals"** nous donne les variables locales qui sont accessibles **à cet endroit du code**, comme le montre ce second exemple qui se concentre sur `locals` à différents points d'une même fonction.

```
import env_locals
```

```
# le code de ce module
show_module(env_locals)
```

Fichier `E:python_code\flotpython-master\w4env_locals.py`

```
-----
| # -- coding: utf-8 --
|
| """
| un module pour illustrer la fonction locals
| """
|
| # pour afficher
| from env_locals_globals import display_env
```



```

|
| def temoin(x):
|     "la fonction témoin"
|     y = x ** 2
|     print(20*'- ', 'locals - entrée:')
|     display_env(locals())
|
|     for i in (1,):
|         for j in (1,):
|             print(20*'- ', 'locals - boucles for:')
|             display_env(locals())
|
|

```

```
env_locals.temoin(10)
```

```

----- locals - entrée:
          x → int
          y → int
----- locals - boucles for:
          i → int
          j → int
          x → int
          y → int

```

83.1.3 Usage pour le formatage de chaînes

Ces deux fonctions ne sont pas d'une utilisation très fréquente. Elles peuvent cependant être utiles dans le contexte du formatage de chaînes, comme on peut le voir dans les deux exemples ci-dessous.

NOTE : si vous pouvez être sûr de disposer d'une version au moins supérieure à python-3.6, alors les *f-strings* vous donnent une alternative bien plus élégante que ce que nous allons voir.

Avec format

On peut utiliser `format` qui s'attend à quelque chose comme :

```
"{nom}".format(nom="Dupont")
```

```
'Dupont'
```

que l'on peut obtenir de manière équivalente, avec le passage d'arguments en `**`, comme on l'a vu en fin de semaine passée :

```
"{nom}".format(**{'nom': 'Dupont'})
```

```
'Dupont'
```

En versant la fonction `locals` dans cette formule on obtient une forme relativement élégante

```

def format_et_locals(nom, prenom, civilite, telephone):
    return "{civilite} {prenom} {nom} : Poste {telephone}".format(**locals())

format_et_locals('Dupont', 'Jean', 'Mr', '7748')

```

```
'Mr Jean Dupont : Poste 7748'
```

Avec l'opérateur %

De manière similaire, avec l'opérateur % - dont nous rappelons qu'il est obsolète - on peut écrire

```
def pourcent_et_locals(nom, prenom, civilite, telephone):  
    return "%(civilite)s %(prenom)s %(nom)s : Poste %(telephone)s"%locals()  
  
pourcent_et_locals('Dupont', 'Jean', 'Mr', '7748')
```

```
'Mr Jean Dupont : Poste 7748'
```

Avec un f-string

Pour rappel si vous disposez de python 3.6, vous pouvez alors écrire simplement - et sans avoir recours, donc, à `locals()` ou autre :

```
# attention ceci nécessite python-3.6  
def avec_f_string(nom, prenom, civilite, telephone):  
    return f"{civilite} {prenom} {nom} : Poste {telephone}"  
  
avec_f_string('Dupont', 'Jean', 'Mr', '7748')
```

```
'Mr Jean Dupont : Poste 7748'
```

Passage d'arguments

84.1 Complément - niveau intermédiaire

84.1.1 Motivation

Jusqu'ici nous avons développé le modèle simple qu'on trouve dans tous les langages de programmation, à savoir qu'une fonction a un nombre fixe, supposé connu, d'arguments. Ce modèle a cependant quelques limitations ; les mécanismes de passage d'arguments que propose python, et que nous venons de voir dans les vidéos, visent à lever ces limitations.

Voyons de quelles limitations il s'agit.

84.1.2 Nombre d'arguments non connu à l'avance

Ou encore : introduction à la forme `*arguments`

Pour prendre un exemple aussi simple que possible, considérons la fonction `print`, qui nous l'avons vu, accepte un nombre quelconque d'arguments.

```
print("la fonction", "print", "peut", "prendre", "plein", "d'arguments")
```

```
la fonction print peut prendre plein d'arguments
```

Imaginons maintenant que nous voulons implémenter une variante de `print`, c'est-à-dire une fonction `error`, qui se comporte exactement comme `print` sauf qu'elle ajoute en début de ligne une balise `ERROR`.

Se posent alors deux problèmes : * D'une part il nous faut un moyen de spécifier que notre fonction prend un nombre quelconque d'arguments.

— D'autre part il faut une syntaxe pour repasser tous ces arguments à la fonction `print`.

On peut faire tout cela avec la notation en `*` comme ceci :

```
# accepter n'importe quel nombre d'arguments
def error(*print_args):
    # et les repasser à l'identique à print en plus de la balise
    print('ERROR', *print_args)

# on peut alors l'utiliser comme ceci
error("problème", "dans", "la", "fonction", "foo")
# ou même sans argument
error()
```

```
ERROR problème dans la fonction foo
ERROR
```

84.1.3 Légère variation

Pour sophistiquer un peu cet exemple, on veut maintenant imposer à la fonction erreur qu'elle reçoive un argument obligatoire de type entier qui représente un code d'erreur, plus à nouveau un nombre quelconque d'arguments pour print.

Pour cela, on peut définir une signature (les paramètres de la fonction) qui

- prévoit un argument traditionnel en première position, qui sera obligatoire lors de l'appel,
- et le tuple des arguments pour print, comme ceci :

```
# le premier argument est obligatoire
def error1(error_code, *print_args):
    message = f"message d'erreur code {error_code}"
    print("ERROR", message, '--', *print_args)

# que l'on peut à présent appeler comme ceci
error1(100, "un", "petit souci avec", [1, 2, 3])
```

```
ERROR message d'erreur code 100 -- un petit souci avec [1, 2, 3]
```

Remarquons que maintenant la fonction `error1` ne peut plus être appelée sans argument, puisqu'on a mentionné un paramètre **obligatoire** `error_code`.

84.1.4 Ajout de fonctionnalités

Ou encore : la forme `argument=valeur_par_defaut`

Nous envisageons à présent le cas - tout à fait indépendant de ce qui précède - où vous avez écrit une librairie graphique, dans laquelle vous exposez une fonction `ligne` définie comme suit. Évidemment pour garder le code simple, nous imprimons seulement les coordonnées du segment :

```
# première version de l'interface pour dessiner une ligne
def ligne(x1, y1, x2, y2):
    "dessine la ligne (x1, y1) -> (x2, y2)"
    # restons simple
    print(f"la ligne ({x1}, {y1}) -> ({x2}, {y2})")
```

Vous publiez cette librairie en version 1, vous avez des utilisateurs ; et quelque temps plus tard vous écrivez une version 2 qui prend en compte la couleur. Ce qui vous conduit à ajouter un paramètre pour `ligne`.

Si vous le faites en déclarant

```
def ligne(x1, y1, x2, y2, couleur):
    ...
```

alors tous les utilisateurs de la version 1 vont **devoir changer leur code** - pour rester à fonctionnalité égale - en ajoutant un cinquième argument 'noir' à leurs appels à ligne.

Vous pouvez éviter cet inconvénient en définissant la deuxième version de ligne comme ceci :

```
# deuxième version de l'interface pour dessiner une ligne
def ligne(x1, y1, x2, y2, couleur="noir"):
    "dessine la ligne (x1, y1) -> (x2, y2) dans la couleur spécifiée"
    # restons simple
    print(f"la ligne ({x1}, {y1}) -> ({x2}, {y2}) en {couleur}")
```

Avec cette nouvelle définition, on peut aussi bien

```
# faire fonctionner du vieux code sans le modifier
ligne(0, 0, 100, 100)
# ou bien tirer profit du nouveau trait
ligne(0, 100, 100, 0, 'rouge')
```

```
la ligne (0, 0) -> (100, 100) en noir
la ligne (0, 100) -> (100, 0) en rouge
```

Les paramètres par défaut sont très utiles

Notez bien que ce genre de situation peut tout aussi bien se produire sans que vous ne publiiez de librairie, à l'intérieur d'une seule application. Par exemple, vous pouvez être amené à ajouter un argument à une fonction parce qu'elle doit faire face à de nouvelles situations imprévues, et que vous n'avez pas le temps de modifier tout le code.

Ou encore plus simplement, vous pouvez choisir d'utiliser ce passage de paramètres dès le début de la conception ; une fonction `ligne` réaliste présentera une interface qui précise : les points concernés, la couleur du trait, l'épaisseur du trait, le style du trait, le niveau de transparence, etc. Il n'est vraiment pas utile que tous les appels à `ligne` reprécisent tout ceci intégralement, aussi une bonne partie de ces paramètres seront très constructivement déclarés avec une valeur par défaut.

84.2 Complément - niveau avancé

84.2.1 Écrire un wrapper

Ou encore : la forme ****keywords**

La notion de *wrapper* - emballage, en anglais - est très répandue en informatique, et consiste, à partir d'un morceau de code souche existant (fonction ou classe) à définir une variante qui se comporte comme la souche, mais avec quelques légères différences.

La fonction `error` était déjà un premier exemple de *wrapper*. Maintenant nous voulons définir un *wrapper* `ligne_rouge`, qui sous-traite à la fonction `ligne` mais toujours avec la couleur rouge.

Maintenant que l'on a injecté la notion de paramètre par défaut dans le système de signature des fonctions, se repose la question de savoir comment passer à l'identique les arguments de `ligne_rouge` à `ligne`.

Évidemment, une première option consiste à regarder la signature de `ligne` :

```
def ligne(x1, y1, x2, y2, couleur="noir")
```

Et à en déduire une implémentation de `ligne_rouge` comme ceci

```
# la version naïve - non conseillée - de ligne_rouge
def ligne_rouge(x1, y1, x2, y2):
    return ligne(x1, y1, x2, y2, couleur='rouge')

ligne_rouge(0, 0, 100, 100)
```

```
la ligne (0, 0) -> (100, 100) en rouge
```

Toutefois, avec cette implémentation, si la signature de `ligne` venait à changer, on serait vraisemblablement amené à changer **aussi** celle de `ligne_rouge`, sauf à perdre en fonctionnalité. Imaginons en effet que `ligne` devienne dans une version suivante

```
# on ajoute encore une fonctionnalité à la fonction ligne
def ligne(x1, y1, x2, y2, couleur="noir", epaisseur=2):
    print(f"la ligne ({x1}, {y1}) -> ({x2}, {y2}) "
          f" en {couleur} - ep. {epaisseur}")
```

Alors le wrapper ne nous permet plus de profiter de la nouvelle fonctionnalité. De manière générale, on cherche au maximum à se prémunir contre de telles dépendances. Aussi, il est de beaucoup préférable d'implémenter `ligne_rouge` comme suit, où vous remarquerez que **la seule hypothèse** faite sur `ligne` est qu'elle accepte un argument nommé `couleur`.

```
def ligne_rouge(*arguments, **keywords):
    # c'est le seul endroit où on fait une hypothèse sur la fonction `ligne`
    # qui est qu'elle accepte un argument nommé 'couleur'
    keywords['couleur'] = "rouge"
    return ligne(*arguments, **keywords)
```

Ce qui permet maintenant de faire

```
ligne_rouge(0, 100, 100, 0, epaisseur=4)
```

```
la ligne (0, 100) -> (100, 0) en rouge - ep. 4
```

84.2.2 Pour en savoir plus - la forme générale

Une fois assimilé ce qui précède, vous avez de quoi comprendre une énorme majorité (99% au moins) du code python.

Dans le cas général, il est possible de combiner les 4 formes d'arguments :

- des expressions (cas des arguments "normaux", dits positionnels)
- des expressions nommées, comme `nom=<expression>`
- des `*expressions`
- des `**expressions`

Vous pouvez [vous reporter à cette page](#) pour une description détaillée de ce cas général.

À l'appel d'une fonction, il faut résoudre les arguments, c'est-à-dire associer une valeur à chaque paramètre formel (ceux qui apparaissent dans le `def`) à partir des valeurs figurant dans l'appel.

L'idée est que pour faire cela, les arguments de l'appel ne sont pas pris dans l'ordre où ils apparaissent, mais les arguments positionnels sont utilisés en premier. La logique est que, naturellement les arguments positionnels (ou ceux qui proviennent d'une `*expression`) viennent sans nom, et donc ne peuvent pas être utilisés pour résoudre des arguments nommés.

Voici un tout petit exemple pour vous donner une idée de la complexité de ce mécanisme lorsqu'on mélange toutes les 4 formes d'arguments à l'appel de la fonction (alors qu'on a défini la fonction avec 4 paramètres positionnels)

```
# une fonction qui prend 4 paramètres simples
def foo(a, b, c, d):
    print(a, b, c, d)
```

```
# on peut l'appeler par exemple comme ceci
foo(1, c=3, *(2,), **{'d':4})
```

```
1 2 3 4
```

```
# mais pas comme cela
try:
    foo (1, b=3, *(2,), **{'d':4})
except Exception as e:
    print(f"OOPS, {type(e)}, {e}")
```

```
OOPS, <class 'TypeError'>, foo() got multiple values for argument 'b'
```

Si le problème ne vous semble pas clair, vous pouvez regarder la [documentation python](#) décrivant ce problème.

85.1 Complément - niveau basique

85.1.1 N'utilisez pas d'objet mutable pour les valeurs par défaut

En python il existe un piège dans lequel il est très facile de tomber. Aussi si vous voulez aller à l'essentiel : **n'utilisez pas d'objet mutable pour les valeurs par défaut.**

Si vous avez besoin d'écrire une fonction qui prend en argument par défaut une liste ou un dictionnaire vide, voici comment faire

```
# ne faites SURTOUT PAS ça
def ne_faites_pas_ca(options={}):
    "faire quelque chose"
```

```
# mais plutôt comme ceci
def mais_plutot_ceci(options=None):
    if options is None:
        options = {}
    "faire quelque chose"
```

85.2 Complément - niveau intermédiaire

85.2.1 Que se passe-t-il si on le fait ?

Considérons le cas relativement simple d'une fonction qui calcule une valeur - ici un entier aléatoire entre 0 et 10 -, et l'ajoute à une liste passée par l'appelant.

Et pour rendre la vie de l'appelant plus facile, on se dit qu'il peut être utile de faire en sorte que si l'appelant n'a pas de liste sous la main, on va créer pour lui une liste vide. Et pour ça on fait :

```
import random

# l'intention ici est que si l'appelant ne fournit pas
# la liste en entrée, on crée pour lui une liste vide
def ajouter_un_aleatoire(resultats=[]):
```

```
resultats.append(random.randint(0, 10))  
return resultats
```

Si on appelle cette fonction une première fois, tout semble bien aller

```
ajouter_un_aleatoire()
```

```
[7]
```

Sauf que, si on appelle la fonction une deuxième fois, on a une surprise !

```
ajouter_un_aleatoire()
```

```
[7, 0]
```

85.2.2 Pourquoi ?

Le problème ici est qu'une valeur par défaut - ici l'expression `[]` - est évaluée **une fois** au moment de la **définition** de la fonction.

Toutes les fois où la fonction est appelée avec cet argument manquant, on va utiliser comme valeur par défaut **le même objet**, qui la première fois est bien une liste vide, mais qui se fait modifier par le premier appel.

Si bien que la deuxième fois on réutilise la même liste **qui n'est plus vide**. Pour aller plus loin, vous pouvez regarder la documentation python sur [ce problème](#).

Licence CC BY-NC-ND Thierry Parmentelat & Arnaud Legout

Arguments *keyword-only*

86.1 Complément - niveau intermédiaire

86.1.1 Rappel

Nous avons vu dans un précédent complément les 4 familles de paramètres qu'on peut déclarer dans une fonction :

1. paramètres positionnels (usuels)
2. paramètres nommés (forme *name=default*)
3. paramètres ***args* qui attrape dans un tuple le reliquat des arguments positionnels
4. paramètres ***kwargs* qui attrape dans un dictionnaire le reliquat des arguments nommés

Pour rappel :

```
# une fonction qui combine les différents  
# types de paramètres  
def foo(a, b=100, *args, **kwargs):  
    print(f"a={a}, b={b}, args={args}, kwargs={kwargs}")
```

```
foo(1)
```

```
a=1, b=100, args=(), kwargs={}
```

```
foo(1, 2)
```

```
a=1, b=2, args=(), kwargs={}
```

```
foo(1, 2, 3)
```

```
a=1, b=2, args=(3,), kwargs={}
```

```
foo(1, 2, 3, bar=1000)
```

```
a=1, b=2, args=(3,), kwargs={'bar': 1000}
```

86.1.2 Un seul paramètre attrape-tout

Notez également que, de bon sens, on ne peut déclarer qu'un seul paramètre de chacune des formes d'attrape-tout ; on ne peut pas par exemple déclarer

```
# c'est illégal de faire ceci
def foo(*args1, *args2):
    pass
```

car évidemment on ne saurait pas décider de ce qui va dans `args1` et ce qui va dans `args2`.

86.1.3 Ordre des déclarations

L'ordre dans lequel sont déclarés les différents types de paramètres d'une fonction est imposé par le langage. Ce que vous avez peut-être en tête si vous avez appris **python2**, c'est qu'à l'époque on devait impérativement les déclarer dans cet ordre :

positionnels, nommés, forme `*`, forme `**`

comme dans notre fonction `foo`.

Ça reste une bonne approximation, mais depuis python-3, les choses ont un petit peu changé suite à l'adoption du [PEP 3102](#), qui vise à introduire la notion de paramètre qu'il faut impérativement nommer lors de l'appel (en anglais : *keyword-only argument*)

Pour résumer, il est maintenant possible de déclarer des **paramètres nommés après la forme `***`**

Voyons cela sur un exemple

```
# on peut déclarer un paramètre nommé **après** l'attrape-tout *args
def bar(a, *args, b=100, **kws):
    print(f"a={a}, b={b}, args={args}, kws={kws}")
```

L'effet de cette déclaration est que, si je veux passer un argument au paramètre `b`, **je dois le nommer**

```
# je peux toujours faire ceci
bar(1)
```

```
a=1, b=100, args=(), kws={}
```

```
# mais si je fais ceci l'argument 2 va aller dans args
bar(1, 2)
```

```
a=1, b=100, args=(2,), kws={}
```

```
# pour passer b=2, je **dois** nommer mon argument
bar(1, b=2)
```

```
a=1, b=2, args=(), kws={}
```

Ce trait n'est objectivement pas utilisé massivement en python, mais cela peut être utile de le savoir :

- en tant qu'utilisateur d'une librairie, car cela vous impose une certaine façon d'appeler une fonction ;
- en tant que concepteur d'une fonction, car cela vous permet de manifester qu'un paramètre optionnel joue un rôle particulier.

Passage d'arguments

87.1 Exercice - niveau basique

```
# pour charger l'exercice
from corrections.exo_distance import exo_distance
```

Vous devez écrire une fonction `distance` qui prend un nombre quelconque d'arguments numériques non complexes, et qui retourne la racine carrée de la somme des carrés des arguments.

Plus précisément : $distance(x_1, \dots, x_n) = \sqrt{\sum x_i^2}$

Par convention on fixe que `distance() = 0`

```
# des exemples
exo_distance.example()
```

| Appel Résultat Attendu | |
|--------------------------|--------------------|
| distance() 0.0 | |
| distance(1) | 1.0 |
| distance(1, 1) | 1.4142135623730951 |

```
# ATTENTION vous devez aussi définir les arguments de la fonction
def distance(votre, signature):
    return "votre code"
```

```
# la correction
exo_distance.correction(distance)
```

87.2 Exercice - niveau intermédiaire

```
# Pour charger l'exercice
from corrections.exo_numbers import exo_numbers
```

On vous demande d'écrire une fonction `numbers`

- qui prend en argument un nombre quelconque d'entiers,
- et qui retourne un tuple contenant
 - la somme
 - le minimum
 - le maximum de ses arguments.

Si aucun argument n'est passé, `numbers` doit renvoyer un tuple contenant 3 entiers 0.

```
# par exemple
exo_numbers.example()
```

| Appel Résultat Attendu | |
|------------------------------------|-------------|
| <code>numbers()</code> (0, 0, 0) | |
| <code>numbers(6)</code> | (6, 6, 6) |
| <code>numbers(10, 7, 5, 6)</code> | (28, 5, 10) |

En guise d'indice, je vous invite à regarder les fonctions *builtin* `sum` <<https://docs.python.org/3/library/functions.html#sum>>, `min` <<https://docs.python.org/3/library/functions.html#min>> et `max` <<https://docs.python.org/3/library/functions.html#max>>.

```
# vous devez définir votre propre signature
def numbers(votre, signature):
    "<votre_code>"
```

```
# pour vérifier votre code
exo_numbers.correction(numbers)
```