



The world of research

Karën Fort

karen.fort@univ-lorraine.fr / <https://members.loria.fr/KFort>

Sources of inspiration

- ▶ Slides from Hee-Soo Choi (PhD student, UL) from a Chiche! presentation (high-school students)

What is research?

What do researchers do?

What is a PhD?



1

Allez sur wooclap.com

2

Entrez le code d'événement dans le bandeau supérieur

 [Activer les réponses par SMS](#)

Code d'événement
USMJCX

<https://app.wooclap.com/USMJCX?from=instruction-slide>

What is research useful for?

should it be useful?

Serendipity: chance and insight

serendipity (*countable* and *uncountable*, plural **serendipities**)

1. The phenomenon of making an **unplanned**, fortunate discovery through a **combination** of unexpected circumstances and **insightful recognition**. [antonyms ▲] [quotations ▼]

Antonyms: **Murphy's law**, **perfect storm**

2. An unsought, unintended or unexpected, but fortunate, discovery or learning experience that occurs by accident.

[synonyms ▲] [quotations ▼]

Synonyms: **chance**, **luck**; *see also* **Thesaurus:luck**

3. The occurrence and development of events by chance in a happy or beneficial way.

<https://en.wiktionary.org/wiki/serendipity#English>

What is research?

What do researchers do?

What is a PhD?



1 Allez sur wooclap.com

2 Entrez le code d'événement dans le bandeau supérieur

 [Activer les réponses par SMS](#)

Code d'événement
USMJCX

<https://app.wooclap.com/USMJCX?from=instruction-slide>

In France, two types of researchers

- ▶ Teachers-researchers (*enseignant·e·s-chercheur·e·s*): do research AND teach classes at the University (192 hours equivalent lab/year)
- ▶ Researchers: do research (don't have to teach at all, but can)

In France, two types of researchers

- ▶ Teachers-researchers (*enseignant·e·s-chercheur·e·s*):
 - ▶ Maître·sse·s de Conférences (MCF): rank B ([Salary \(traitement\)](#))
 - ▶ Professeur·e·s des Universités (PU): rank A ([Salary \(traitement\)](#))
- ▶ Researchers:
 - ▶ Chargé·e·s de recherche (CR): rank B
 - ▶ Directeur·ice·s de recherche (DR): rank A

CR vs MCF ;-)



Generated from: <https://imgflip.com/memegenerator/116570292/Mel-Gibson-and-Jesus-Christ>

How do you become a researcher in France?

Be MCF or CR ⇒ be selected through a competitive exam (concours)
⇒ have a **PhD** (3 or 4 years of work)
⇒ have a Master degree

How do you become a researcher in France?

Be MCF or CR ⇒ be selected through a competitive exam (concours)
⇒ have a **PhD** (3 or 4 years of work)
⇒ have a Master degree

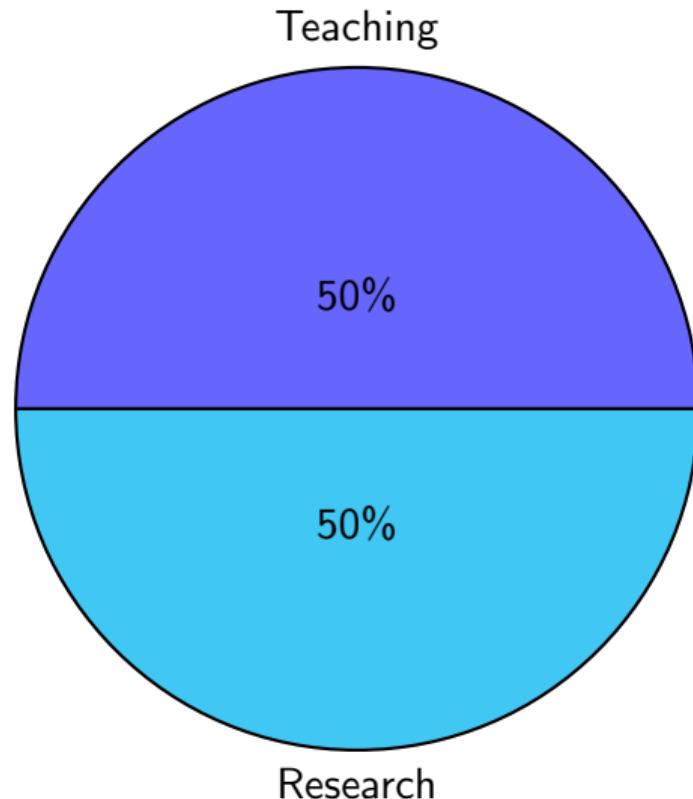
Be PU or DR ⇒ be selected through a competitive exam (concours)
⇒ have a **HDR** (habilitation à diriger des recherches, similar to a PhD)
⇒ have a PhD

Status?

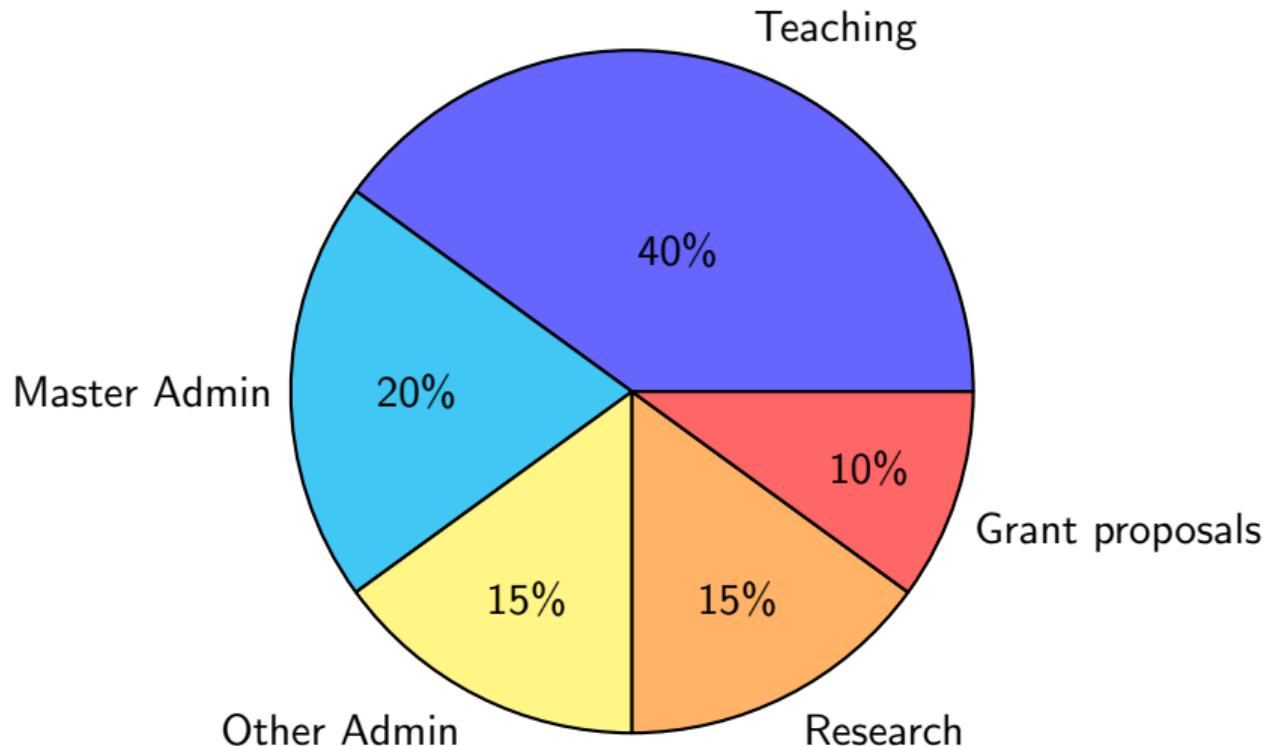
In France, MCF, PU, CR and DR are all [civil servants](#) (fonctionnaires d'Etat)

What about in other countries?

My work (in theory)



My work (in practice)



Presenting a poster at a conference (Mumbai, 2012)



Presenting a poster at a conference (Mumbai, 2012)



Presenting at a conference (Portorož, 2016)





1

Allez sur wooclap.com

2

Entrez le code d'événement dans le bandeau supérieur



Activer les réponses par SMS

Code d'événement
USMJCX

<https://app.wooclap.com/USMJCX?from=instruction-slide>

What is research?

What do researchers do?

What is a PhD?

Doing a PhD in computer science (CS)

- ▶ 3 years (often 4, rarely more)
- ▶ doing research (and a bit of teaching)
- ▶ with your advisor(s)
- ▶ writing and defending a thesis
- ▶ publishing papers during the PhD is **mandatory** (different in other disciplines)

Prejudices about PhDs

Doing a PhD is like doing a Master degree – **No!**

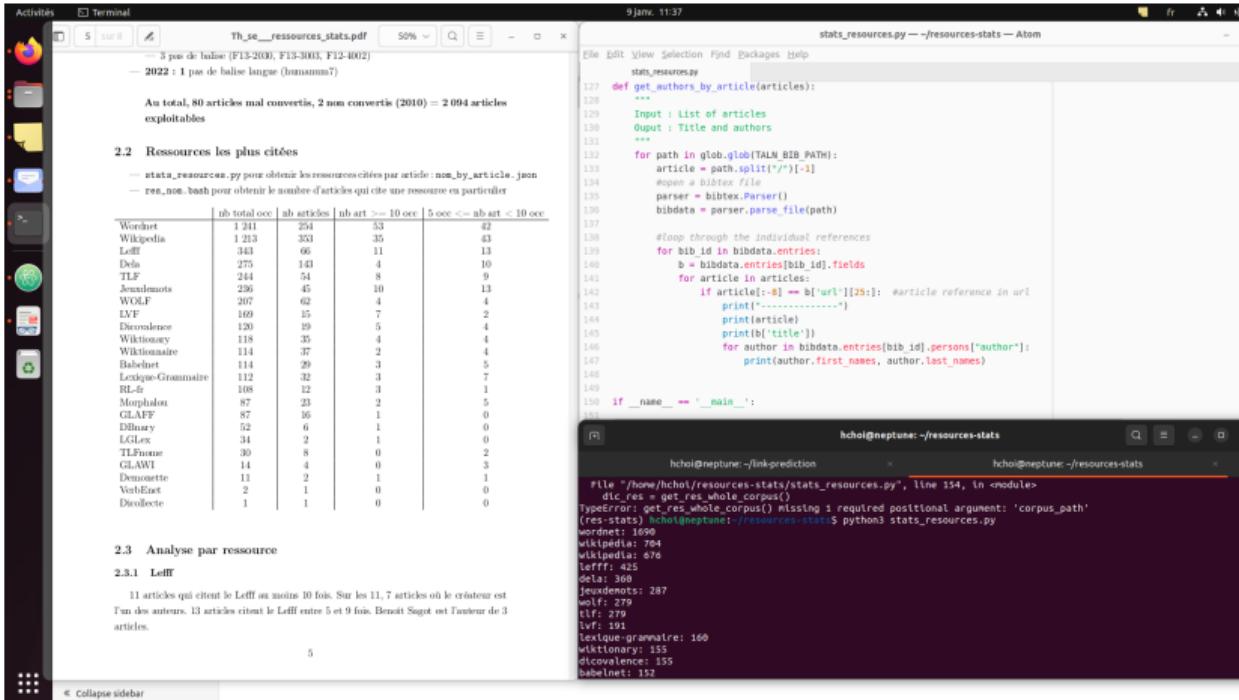
- ▶ No exam, no grades
- ▶ You can/should attend some classes, that you choose, and that are related to your subject
- ▶ Autonomy and freedom
- ▶ New problems

A PhD is not remunerated – **Not in CS**

- ▶ Doctoral Contracts: doctoral school, lab, ANR projects
(+€1,700 net/month, without teaching)
- ▶ CIFRE: PhD in a company and a lab
(salary is now similar)

A PhD student day to day life

Reading papers, running experiments and discussing with your supervisor(s) (in general, once a week)



The screenshot shows a Linux desktop environment with a dark theme. On the left, a vertical sidebar contains icons for various applications. The main window area has three open panes:

- Terminal:** Shows a command-line interface with the following text:

```
hchoi@neptune:~/link-prediction
File "/home/hchoi/resources-stats/stats_resources.py", line 154, in <module>
    dic_res = get_res_whole_corpus()
TypeError: get_res_whole_corpus() missing 1 required positional argument: 'corpus_path'
(hchoi@neptune:~/resources-stats)$ python3 stats_resources.py
wordnet: 1690
wikipedia: 704
wiktionary: 676
lefff: 425
dela: 366
jedopressis: 287
wlf: 279
tlf: 279
lrf: 191
lexique-grammaire: 166
wiktionary: 155
dicivalence: 155
babelnet: 152
```
- Code Editor:** Shows a Python script named `stats_resources.py` with the following content:

```
def get_authors_by_article(articles):
    """
    Input : List of articles
    Output : Title and authors
    """
    for path in glob.glob(TALN_BIB_PATH):
        article = path.split('/')[-1]
        open a bibtex file
        parser = bibtex.Parser()
        bibdata = parser.parse_file(path)

    #loop through the individual references
    for bib_id in bibdata.entries:
        b = bibdata.entries[bib_id].fields
        for article in articles:
            if article[-8] == b['url'][125]: #article reference in url
                print("-----")
                print(article)
                print(b['title'])
                for author in bibdata.entries[bib_id].persons["author"]:
                    print(author.first_names, author.last_names)
```
- PDF Viewer:** Shows a PDF document titled `Th_se_ressources_stats.pdf` with the following content:

3 pas de balise (F13-200, F13-300, F12-4002)
2022 : 1 pas de balise longue (Innsmouth7)

Au total, 80 articles mal convertis, 2 non convertis (2010) = 2 094 articles exploitables

2.2 Ressources les plus citées

stats_resources.py pour obtenir les ressources citées par article : nos_by_article.json
res_noe.txt pour obtenir le nombre d'articles qui cite une ressource en particulier

	nb total occ	nb articles	nb art >= 10 occ	5 occ <= nb art < 10 occ
Wordnet	1 241	254	53	42
Wikipedia	1 213	353	35	43
Lefff	343	66	11	13
Dela	275	143	4	10
TLF	244	54	8	9
Jeudismatin	236	45	10	13
WOLF	207	62	4	4
LRF	169	15	7	2
Dicivalence	120	19	5	4
Wiktionary	118	35	4	4
Wiktionnaire	114	37	2	4
Babelnet	114	29	3	5
Lexique-Grammaire	112	32	3	7
RL-fr	108	12	3	1
Morphitou	87	23	2	5
GLAFF	87	16	1	0
DBnary	52	6	1	0
LG-lex	34	2	1	0
TLFonline	30	8	0	2
GLAWI	14	4	0	3
Democrite	11	2	1	1
VerbiEst	2	1	0	0
Dicolecte	1	1	0	0

2.3 Analyse par ressource

2.3.1 Lefff

11 articles qui citent le Lefff au moins 10 fois. Sur les 11, 7 articles où le créateur est l'un des auteurs, 13 articles citent le Lefff entre 5 et 9 fois. Benoît Sagot est l'auteur de 3 articles.

From Hee-Soo Choi, with her approval

A PhD student day to day life

Writting papers, publishing them and presenting them at conferences



From Hee-Soo Choi, with her approval

A PhD student day to day life

Interacting with other researchers

- ▶ Weekly meeting with advisor(s)
- ▶ Team meeting (once a week)
- ▶ Conferences, seminars
- ▶ Summer schools
- ▶ Doctoral training

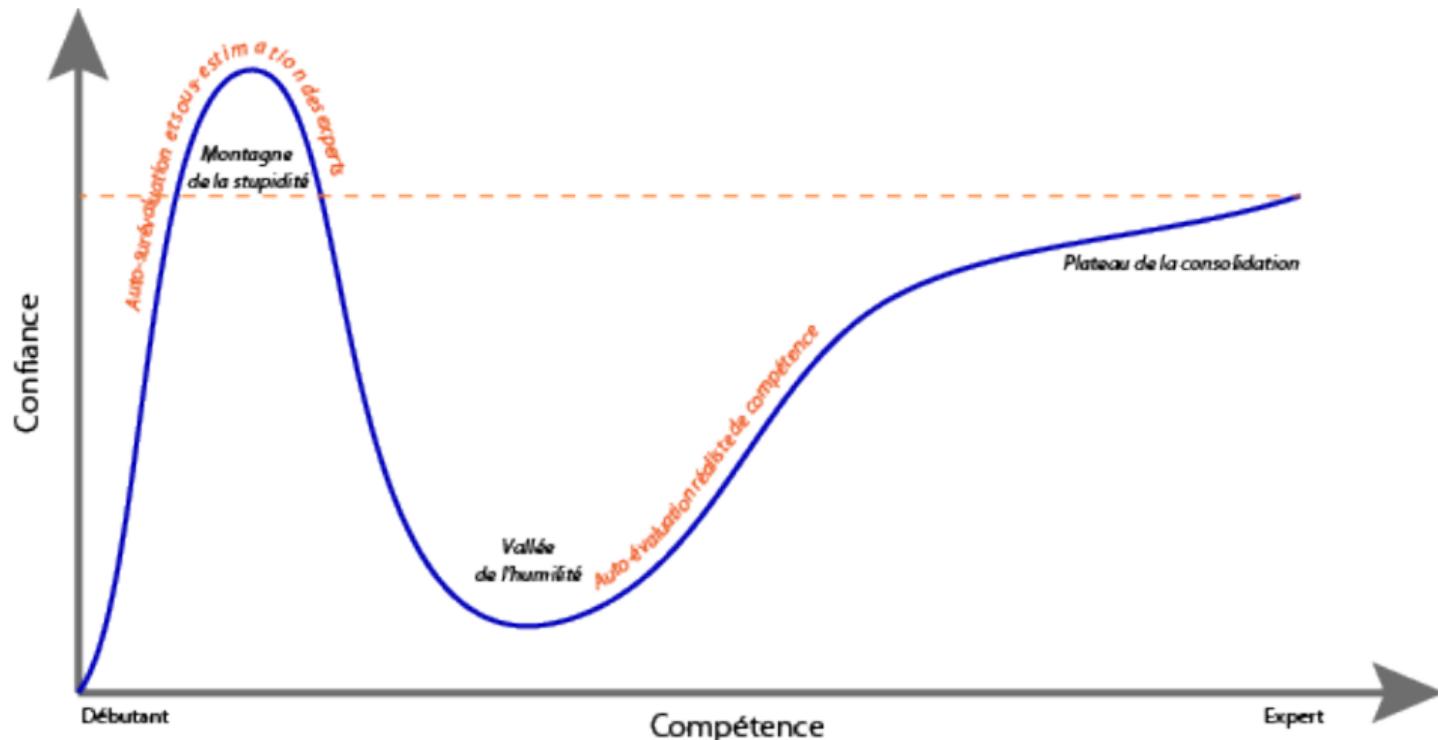
From Hee-Soo Choi, with her approval

A PhD student day to day life

Teaching classes

- ▶ During the 3 years of your contract:
 - ▶ not compulsory
 - ▶ as *vacataire* (lecturer)
 - ▶ paid by the hour (around €43/hour (gross))
- ▶ After the 3 years of PhD contract, as an ATER (Temporary research and teaching assistant):
 - ▶ teach 192 hours + finish your PhD or do post-doctoral research
 - ▶ paid €1,765 (net)

From the Dunning-Kruger effect to the imposter syndrome



Shloren, CC BY-SA 4.0, via Wikimedia Commons

Insights from (some) Semagramme's PhD students

- ▶ Importance of the advisor(s): more important than the subject
- ▶ Importance of the team (support): do not stay isolated
- ▶ Importance of the lab
- ▶ Read and write a lot

Advisor or advisors?

It is possible to have two advisors (even three):

- + different skills, different points of views
- + different relationship (less stress)
- organizational constraints (organizing meetings with more persons)
- disagreements may arise between advisors

PhD students (in NLP) are "young researchers" rather than students

Therefore:

- ▶ they review papers for conferences (with the help of their supervisor)
- ▶ they teach as soon as they start their PhD

Finding a PhD in CS \Rightarrow finding an advisor+grant (+subject)

Different possibilities:

- ▶ Best case scenario (less common): the advisor(s) has the money (through, for ex, an ANR project) \Rightarrow convince them you're the right person
- ▶ The advisor(s) do(es) not have (all) the money (usual case):
 - ▶ you'll need to convince them you're the right person **and**
 - ▶ you'll need to apply for fundings (at the lab or some other level) \Rightarrow competition

Finding a PhD in CS \Rightarrow you don't come up with the subject

... the advisor(s) do(es)

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography
- ▶ evaluated by at least two **reviewers** (*rapporteurs*), who write a report about the thesis and **accept** or **reject** (usually, postpone) the defense

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography
- ▶ evaluated by at least two **reviewers** (*rapporteurs*), who write a report about the thesis and **accept** or **reject** (usually, postpone) the defense
- ▶ if accepted, the thesis is defended in front of a **jury** composed of:

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography
- ▶ evaluated by at least two **reviewers** (*rapporteurs*), who write a report about the thesis and **accept** or **reject** (usually, postpone) the defense
- ▶ if accepted, the thesis is defended in front of a **jury** composed of:
 - ▶ the reviewers

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography
- ▶ evaluated by at least two **reviewers** (*rapporteurs*), who write a report about the thesis and **accept** or **reject** (usually, postpone) the defense
- ▶ if accepted, the thesis is defended in front of a **jury** composed of:
 - ▶ the reviewers
 - ▶ examiners (at least two)

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography
- ▶ evaluated by at least two **reviewers** (*rapporteurs*), who write a report about the thesis and **accept** or **reject** (usually, postpone) the defense
- ▶ if accepted, the thesis is defended in front of a **jury** composed of:
 - ▶ the reviewers
 - ▶ examiners (at least two)
 - ▶ the advisor(s)

The PhD thesis in CS

Esp. in NLP:

- ▶ more than 100 pages, less than 300 (but no limit)
- ▶ state of the art, experiments, results, discussion, bibliography
- ▶ evaluated by at least two **reviewers** (*rapporteurs*), who write a report about the thesis and **accept** or **reject** (usually, postpone) the defense
- ▶ if accepted, the thesis is defended in front of a **jury** composed of:
 - ▶ the reviewers
 - ▶ examiners (at least two)
 - ▶ the advisor(s)
 - ▶ invited members (unofficial advisors)