SH Dr. Soufiane Hourri

drsoufiane.com

MASTER'S PROGRAMME · SECOND YEAR

# Speech Processing

Session 02 · Academic literacy: reading, finding and writing research papers

**Dr. Soufiane Hourri, PhD, HDR**

Associate Professor & AI Consultant

École Supérieure de Technologie de Safi · Cadi Ayyad University

**2026-2027**

# Today's session

01 Why academic literacy matters

02 The anatomy of a research paper

03 Reading a paper in three passes

04 Finding papers fast

05 Academic integrity and citation

06 Workshop: build your own outline

**TODAY'S FOCUS**

Researchers rarely fail because they cannot code.

They fail because they cannot read and write scientifically.

Today is practical: one real paper, one real search, one outline.

# Learning objectives

**By the end of this session, you will be able to**

1. Identify IMRaD and the four parts of an abstract
2. Judge whether a paper is useful in 10 to 15 minutes
3. Search scholarly databases with precise keywords and snowballing
4. Paraphrase and cite sources without plagiarism
5. Outline an introduction with the three moves

**PREREQUISITES**

- None: this session is about method
- A laptop or phone with internet
- An idea of your thesis topic, even vague

01

CHAPTER

# Why academic literacy matters

What goes wrong when students write their first research report, and what we mean by academic literacy.

# Why master's students struggle

- Copying the structure of random blogs or tutorials
- Missing the conventions of academic writing
- An unclear problem statement: the reader cannot tell what is solved
- A literature review that lists papers without connecting them
- Results without comparison, and claims without evidence

**KEY TAKEAWAY**

Good research writing is learnable: it follows a small number of patterns that you can copy.

# What is academic literacy?

DEFINITION

**Academic literacy is the ability to find, read, evaluate, use and produce academic texts in your field.**

EXAMPLE

Before proposing a Darija speech recognizer, you must find what exists, read it critically, cite it correctly and explain clearly what you add.

FIVE ACTIONS

**Find · Read · Evaluate · Use · Produce**

Today: one exercise for each

# Quick reflection · 2 minutes

- When you open a research paper, what makes you close it?
  - Too long or too dense?
  - Too much mathematics without explanation?
  - Unfamiliar vocabulary?
  - You do not know where to start?

**THE CURE**

A reading strategy. You do not read a paper from the first line to the last.

# Where papers fit in a research project

**01**

### Question

What problem do you want to solve?

**02**

### Literature

What is already known? Where is the gap?

**03**

### Method

Data, model, experiments, metrics

**04**

### Results

What did you find? How does it compare?

**05**

### Writing

Report, thesis or paper

02

CHAPTER

# The anatomy of a research paper

Every research paper follows the same funnel: from a broad context to a specific contribution, and back.

# IMRaD: the standard structure

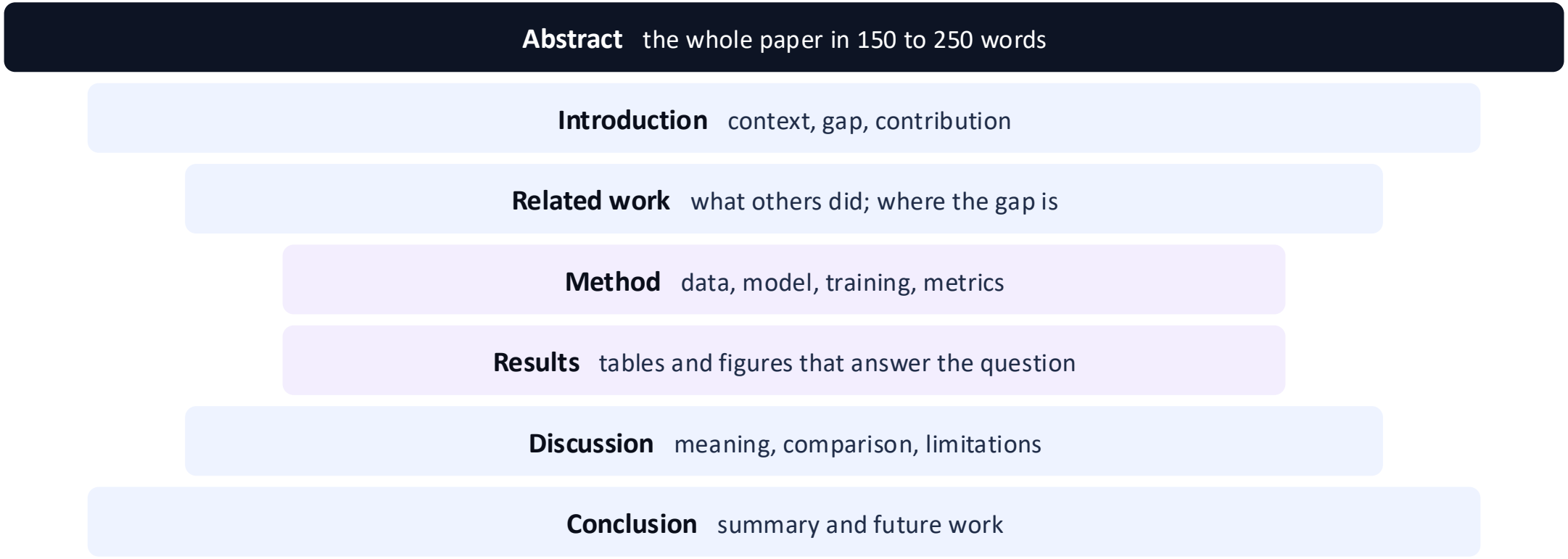


Figure 2.1. IMRaD (Introduction, Methods, Results and Discussion), with the abstract before and the conclusion after. Broad at the start and the end, specific in the middle.

# What each part must answer

## Before the results

**Abstract**

Why should I read this?

**Introduction**

What problem, which gap, what contribution?

## The core

**Method**

Could someone repeat it?

**Results**

What did you measure, against what?

## After the results

**Discussion**

What does it mean? What are the limits?

**Conclusion**

What should the reader remember?

# The abstract in four parts

**01**

### Problem

"This paper addresses the challenge of..."

**02**

### Method

"We propose / adopt / apply..."

**03**

### Findings

"Experiments on [dataset] show that..." with numbers

**04**

### Conclusion

"These results suggest..."

**+**

### Length

150 to 250 words, no citations, no undefined acronyms

# Exercise · Find the four parts of an abstract

**GOAL**

Recognise problem, method, findings and conclusion in a real abstract.

**TIME**

**10 min, in pairs**

**INSTRUCTIONS**

1. Open the abstract of the Whisper paper (Radford et al., 2022, arXiv:2212.04356)
2. Underline the problem in one colour, the method in another
3. Circle every number: these are the findings
4. Find the sentence that gives the conclusion
5. Is any of the four parts missing or weak?

**DELIVERABLE**

The annotated abstract, shared with the class.

# The introduction in three moves

THE CARS MODEL

**An introduction creates a research space in three moves: establish a territory, establish a niche (the gap), then occupy the niche (Swales, 1990).**

KEY INSIGHT

If your introduction contains these three moves, in this order, it already reads as academic writing.

MOVE 1 · TERRITORY

Why the field matters

MOVE 2 · NICHE

What is missing

MOVE 3 · OCCUPATION

What this paper does

# Typical sentences for each move

| Move | Purpose | Typical sentences |
|---|---|---|
| 1 · Territory | Show importance and trends | "In the past decade, much research has focused on..." |
| | | "Speech recognition has become central to..." |
| 2 · Niche | Reveal the gap | "However, few studies have addressed..." |
| | | "It remains unclear whether..." |
| 3 · Occupation | State the contribution | "The purpose of this study is to..." / "We propose..." |
| | | "The rest of the paper is organised as follows..." |

Table 2.1. Phrase bank for the three moves, adapted from Swales (1990). Use them as patterns, not as text to copy.

# Related work and method

## Related work

- Organise by themes, not by authors
- Compare and connect the papers
- Say what each group of papers cannot do
- End with the gap that your method fills

## Method

- Data: source, size, language, split
- Preprocessing: sample rate, features, cleaning
- Model and training setup
- Evaluation metrics and baselines
- Enough detail for someone to repeat it

# Results, discussion and conclusion

## Results

- Only what answers the research question
- Tables and figures, each mentioned in the text
- Compare with baselines and earlier work
- Report numbers, not impressions

## Discussion and conclusion

- Remind the main findings
- Explain them and compare with the literature
- State the limitations honestly
- Conclude: what to remember, what comes next

# Knowledge check

QUESTION

**"However, no publicly available dataset exists for Darija emotional speech." Where does this sentence belong?**

A

Abstract, findings

B

Introduction, move 2 (the niche)

C

Method, data

D

Conclusion, future work

# 03

CHAPTER

# Reading a paper in three passes

Stop reading line by line. Decide quickly whether a paper is useful, then go deeper only when it is.

# The three-pass method

**01**

**Pass 1 · 5 min**

Title, abstract, headings, figures, conclusion

**02**

**Decide**

In my topic? Recent? Credible venue?

**03**

**Pass 2 · 10 to 60 min**

Introduction, method, results; note problem, gap, data, result

**04**

**Decide**

Can I use it, cite it, compare with it?

**05**

**Pass 3 · hours**

Details, equations, experiments; try to reproduce

# What to look at in pass 1

- Title and abstract: is the problem close to mine?
- Venue and year: journal, conference, or preprint? Recent enough?
- Figures and tables: what is measured, on which dataset?
- Conclusion: what do the authors claim?
- References: do I recognise the key papers of my field?

**KEY TAKEAWAY**

Pass 1 is a filter, not reading. Most papers stop here.

# Activity · Extract the key information

**GOAL**

Apply pass 1 and pass 2 to the Whisper paper and extract five items.

**TIME**

**15 min, in pairs**

**INSTRUCTIONS**

1. Pass 1 (5 min): title, abstract, headings, figures, conclusion
2. Pass 2 (10 min): introduction and the main results table
3. Fill the table on the next slide, in your own words
4. Give the section where you found each item

**DELIVERABLE**

The completed table, one per pair.

# Extraction table

| Item | In your own words | Section |
|---|---|---|
| Problem | | |
| Gap | | |
| Contribution | | |
| Method (one line) | | |
| Main result | | |

Table 3.1. Copy this table for every paper you read in pass 2. It becomes the raw material of your related work section.

# Highlighting everything

✗ INCORRECT APPROACH

Read from the first line, highlight half the page, take no notes

A week later, nothing is left

✓ CORRECTION

Read with a question in mind

Write five items per paper in a table, in your own words

Store the paper and notes in a reference manager

WHY IT FAILS

Highlighting feels like work but stores nothing. Writing in your own words is what you will reuse when you write.

# Break · 15 minutes

After the break: finding papers fast, and citing them correctly.

04

CHAPTER

# Finding papers fast

Where to search, how to build a search string, and how to follow references in both directions.

# Where to search

| Source | Strength | Be careful |
|---|---|---|
| Google Scholar | Broadest coverage; "Cited by" links | Mixes peer-reviewed and other documents |
| IEEE Xplore | ICASSP, IEEE/ACM TASLP, engineering | Full text may need university access |
| ACM Digital Library | Computer science, HCI, algorithms | Full text may need university access |
| ISCA Archive | Interspeech papers, free | Speech only |
| arXiv | The latest preprints, free | Not always peer reviewed |
| Semantic Scholar | Summaries and citation graphs | Check details in the original paper |

Table 4.1. Main sources for speech and AI papers. ICASSP and Interspeech are the two main speech conferences.

# Building a search string

- **Topic:** Arabic speech emotion recognition with deep learning
- `"speech emotion recognition" AND arabic`
- `"speech emotion recognition" AND (arabic OR darija)`
- `("low-resource" OR "under-resourced") "speech emotion"`
- `"emotion recognition" wav2vec OR WavLM`

**OPERATORS**

"..." exact phrase · AND narrows · OR widens · ( ) groups · site: limits the domain

# Exercise · Search your topic

**GOAL**

Find five relevant, recent papers on a speech topic with a good search string.

**TIME**

**12 min, individual**

**INSTRUCTIONS**

1. Choose your topic (or: Darija speech recognition)
2. Write the key concepts and two synonyms for each
3. Build a search string with quotes, AND, OR
4. Run it in Google Scholar and in one other source
5. Keep 5 papers from the last 5 years; apply pass 1 to each

**DELIVERABLE**

Your search string and 5 references.

# Snowballing: follow the references

### Backward snowballing

- Start from one good paper
- Read its reference list
- Find the earlier, foundational work
- Useful to understand where an idea comes from

### Forward snowballing

- Start from the same paper
- Click "Cited by" in Google Scholar
- Find the later work that builds on it
- Useful to reach the current state of the art

# Is this source reliable?

- Peer reviewed? Journal or conference paper, or preprint?
- Venue: a recognised conference or journal of the field?
- Evidence: datasets named, metrics given, baselines compared?
- Reproducibility: code or data available?
- Warning signs: vague claims, no comparison, unknown "journals" that publish anything for a fee

**KEY TAKEAWAY**

Cite peer-reviewed work for claims. Use preprints with care, and say that they are preprints.

# Your literature matrix

| Paper | Problem / gap | Method and data | Main result | Why it matters to me |
|---|---|---|---|---|
| [1] Author, year | | | | |
| [2] Author, year | | | | |
| [3] Author, year | | | | |

Table 4.2. One row per paper read in pass 2. Sort the rows by theme: the themes become the subsections of your related work.

# Knowledge check

QUESTION

## You found a 2025 arXiv paper reporting a very low error rate on Darija speech recognition. What should you do?

A

Cite it as an established result

B

Ignore it: preprints are useless

C

Use it, say it is a preprint, and check for a peer-reviewed version

D

Copy its results table into your report

05

CHAPTER

# Academic integrity and citation

Using other people's work honestly: quoting, paraphrasing, citing, and checking what AI tools give you.

# Three ways to use a source

## ✗ Copy-paste

**What**

Original text without quotation marks

**Verdict**

**Plagiarism**

## ✗ Paraphrase, no citation

**What**

Ideas reworded, source not credited

**Verdict**

**Still plagiarism**

## ✓ Paraphrase + citation

**What**

Your own words, source credited

**Verdict**

**Correct academic practice**

# Exercise · Paraphrase and cite

**GOAL**

Turn a source sentence into a correct paraphrase with a citation.

**TIME**

**8 min, individual**

**INSTRUCTIONS**

1. Take one sentence from the abstract of one of your five papers
2. Close the paper; write the idea in your own words
3. Change the structure, not only a few words
4. Add the citation in IEEE style: [1]
5. Swap with your neighbour: is it really a paraphrase?

**DELIVERABLE**

Original sentence, paraphrase, reference.

# Citation styles and tools

## Two common styles

- **IEEE:** numbered, [1], [2]; usual in engineering and speech
- **APA:** author and year, (Smith, 2023)
- Choose one style and use it everywhere
- Every citation in the text appears in the reference list, and the reverse

## Tools

- **Zotero:** free reference manager; saves papers from the browser
- **BibTeX:** reference format used with LaTeX
- **Overleaf:** online LaTeX editor; conference templates
- Export references from Google Scholar or the publisher

# Trusting AI-generated references

✗ INCORRECT APPROACH

Ask a chatbot for "10 references on Darija ASR"

Paste them into your report without checking

✓ CORRECTION

Find every reference in Google Scholar or the publisher

Open it and check that it says what you cite it for

Export the reference from the source, not from the chatbot

WHY IT FAILS

AI tools can produce references that look real but do not exist, or that do not support the claim. A fabricated reference is a serious integrity problem.

06

CHAPTER

# Workshop: build your own outline

Write the skeleton of your research introduction, in reverse order, then get feedback from a classmate.

# Write the introduction in reverse order

**01**

### Occupy the niche

3 to 4 sentences: purpose, method, data, expected result

**02**

### Establish the niche

2 to 3 sentences: the gap your work fills

**03**

### Establish the territory

2 to 3 sentences: why the field matters

**04**

### Read in order

Territory, niche, occupation: check the logic

**05**

### Add citations

Every claim about the field gets a source

# Workshop · Your thesis outline

GOAL

Write a six-part outline for your own research topic.

TIME

**20 min + 10 min feedback**

INSTRUCTIONS

1. 1 Territory: "In the past decade, much research has focused on..."
2. 2 Niche: "However, few studies have... / It remains unclear..."
3. 3 Occupation: "The purpose of this study is to..."
4. 4 Method: "The data used in this study will come from..."
5. 5 Expected results: "We expect that..." · 6 Limitations: "This study may be limited by..."

DELIVERABLE

One page, six parts, at least three citations.

# A worked example: Darija speech recognition

- **Territory:** Neural speech recognition has improved rapidly, and large multilingual models now cover many languages.
- **Niche:** However, Moroccan Darija has little transcribed audio, and code-switching with French remains difficult.
- **Occupation:** We propose to fine-tune a pretrained multilingual model on a small Darija corpus and evaluate it on code-switched speech.
- **Limitations:** Few speakers, one recording condition, a small test set.

YOUR TURN

Each claim above needs a citation. Which papers from your search would you use?

# Exit question

QUESTION

**You have 10 minutes to decide whether a new paper is useful for your thesis. What do you do?**

A

Read it from the first line to the last

B

Pass 1: title, abstract, headings, figures, conclusion

C

Read only the related work section

D

Ask a chatbot to summarise it, and cite the summary

# Recap

## KEY TAKEAWAYS

- IMRaD: a funnel; abstract in four parts
- Introduction: territory, niche, occupation
- Read in three passes; note five items per paper
- Search with quotes, AND, OR; snowball both ways
- Paraphrase, cite, and verify every reference

## HOMEWORK

- Pick one paper in your thesis area
- Extract problem, gap, method, dataset, result
- Write your introduction (8 to 10 lines) with the three moves
- Bring it next session for feedback

# References

(1) Swales, J. M. (1990). Genre Analysis: English in Academic and Research Settings. Cambridge University Press.

(2) Keshav, S. (2007). How to read a paper. ACM SIGCOMM Computer Communication Review, 37(3), 83–84.

(3) Radford, A., et al. (2022). Robust speech recognition via large-scale weak supervision. arXiv:2212.04356

(4) Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. EASE 2014.

(5) IEEE. IEEE Reference Guide. ieeeauthorcenter.ieee.org

(6) Zotero. zotero.org · Overleaf. overleaf.com · ISCA Archive. isca-archive.org

SH

# Thank you · Questions?

Next session: bring your introduction (8 to 10 lines) for feedback.

**Dr. Soufiane Hourri, PhD, HDR**

drsoufiane.com