

Text revision in Scientific Writing Assistance: An Overview

 Nantes
Université

 LS2N

 TALN

Léane Jourdan,
Florian Boudin,
Richard Dufour,
Nicolas Hernandez

Contact: leane.jourdan@ls2n.fr



ECIR
2023

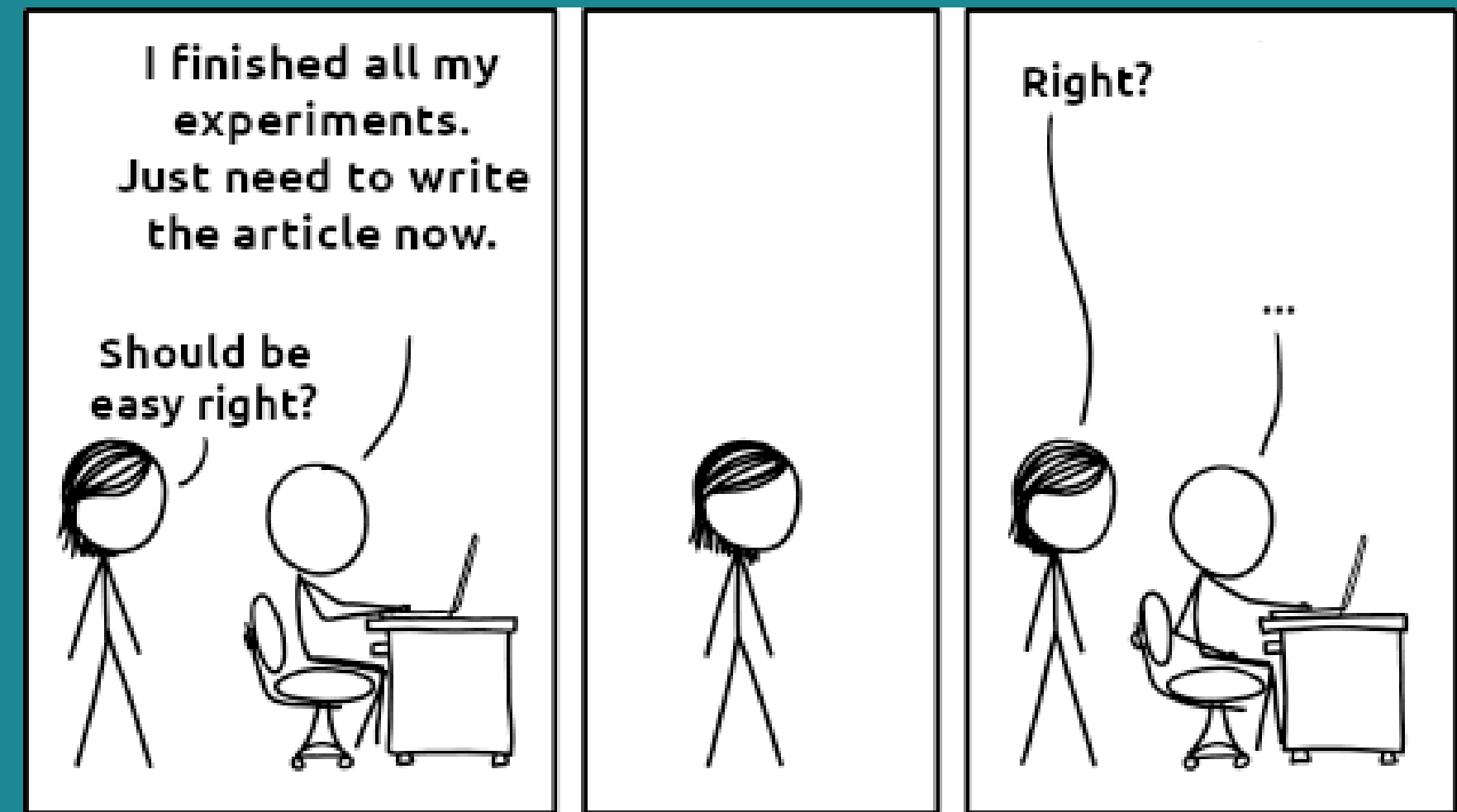
BIR 2023

13th International Workshop on
Bibliometric-enhanced Information Retrieval

Introduction

Context:

- Scientific writing assistance tools
- Focus on text revision



©xkcd

Motivations:

- Writing an article is challenging
- Strong writing skills are essential
- Especially difficult for junior researchers and non-native English speakers

Outline

01

Definition of the research article genre and its specificities

02

Definition of the task of text revision in scientific writing assistance

03

Overview of the current approaches and NLP tools for scientific text revision

04

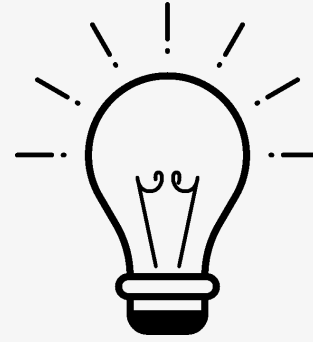
Challenges and future directions

1 – What is Scientific writing?

- Scientific Writing \subset Academic writing
- **Academic writing**
 - **Context** : University setting
 - **Form**: Essay, thesis, syllabus, articles ...
- **Scientific writing**
 - **Form**: articles published in journals or conferences
 - **Specificities**:
 - mostly written in English
 - concise, precise, clear
 - codified usage of tenses, pronouns and terminology



The structure of scientific articles: IMRaD



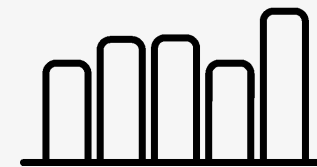
Introduction:

- Gives context and problematic



Methods:

- Describes the research design



Results:

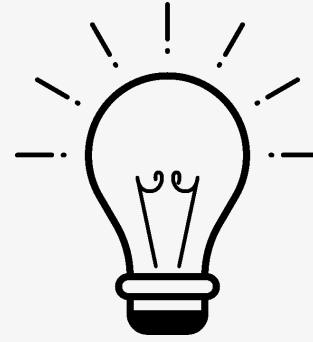
- Presents the findings of the study



Discussion:

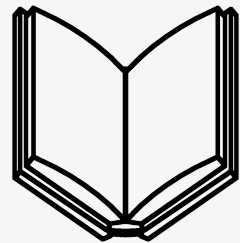
- Interprets and discusses the results
- Suggests directions for future research

The structure of scientific articles: IMRaD



Introduction:

- Gives context and problematic



Literature review/Related work:

- Discusses previous research



Methods:

- Describes the research design



Results:

- Presents the findings of the study



Discussion:

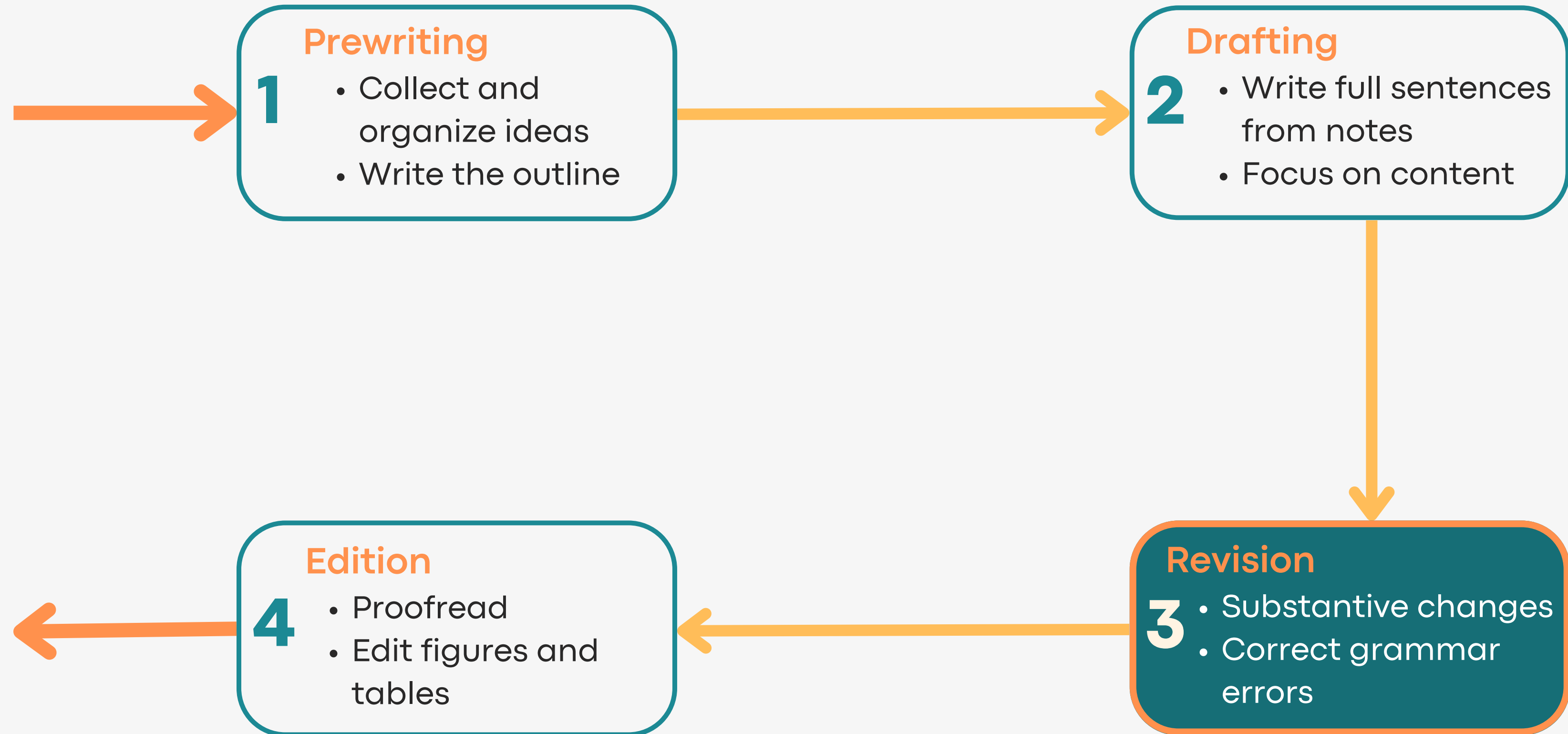
- Interprets and discusses the results
- Suggests directions for future research



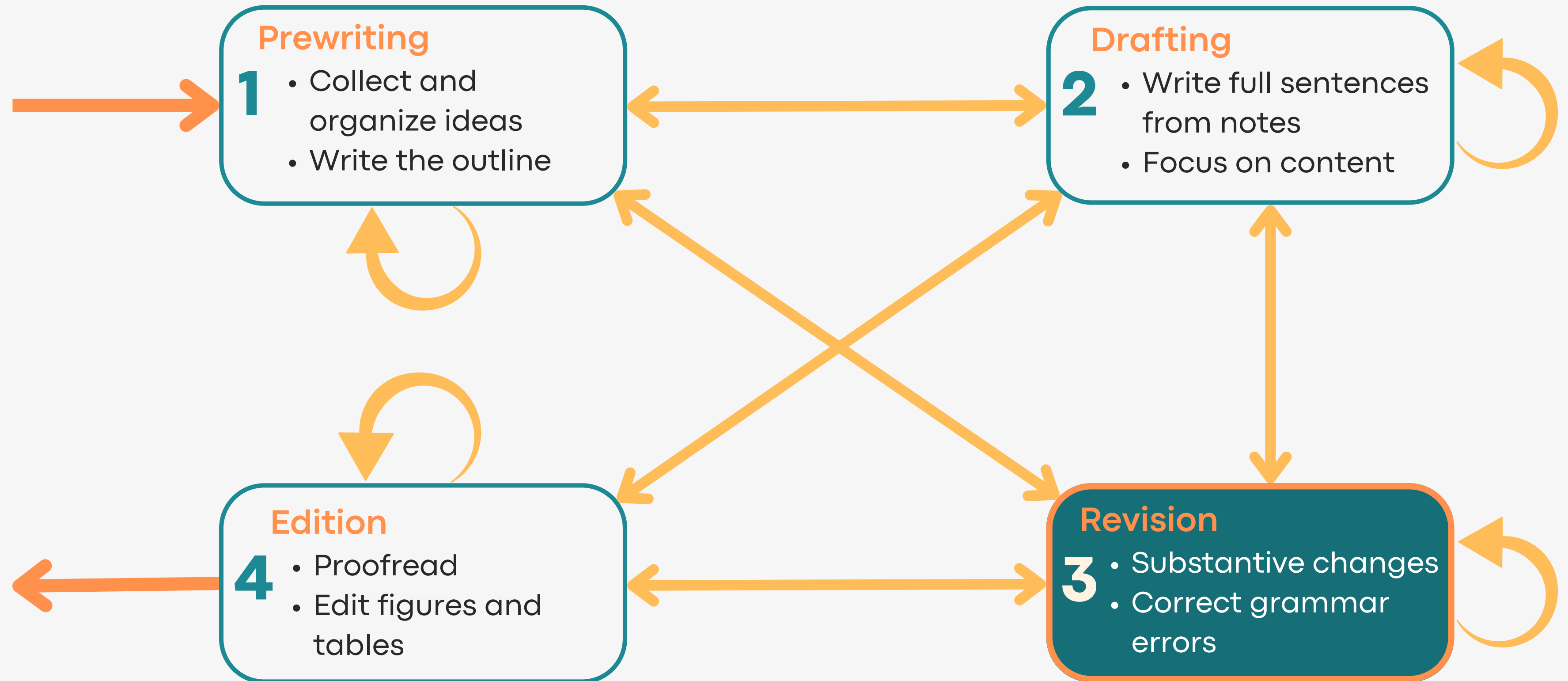
Conclusion:

- Summarizes the key elements

The writing process of scientific articles



The writing process of scientific articles



Modelization of the argumentative structure

The CARS model (Swales, 1990)

Move 1 Establishing a territory

- Step 1 Claiming centrality and/or
- Step 2 Making topic generalization(s) and/or
- Step 3 Reviewing items of previous research.

Move 2 Establishing a niche

- Step 1A Counter claiming or
- Step 1B Indicating a gap or
- Step 1C Question-raising or
- Step 1D Continuing a tradition.

Move 3 Occupying a niche

- Step 1A Outlining purposes or
- Step 1B Announcing present research
- Step 2 Announcing principle findings
- Step 3 Indicating RA structure

Declining rhetorical effort

Weakening Knowledge claims

Increasing explicitness

1. Introduction

The purpose of this paper is to provide a comprehensive overview of the current state of research on the effects of social media on mental health. The paper will focus on the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy. The paper will also discuss the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy.

The paper will focus on the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy. The paper will also discuss the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy.

The paper will focus on the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy. The paper will also discuss the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy.

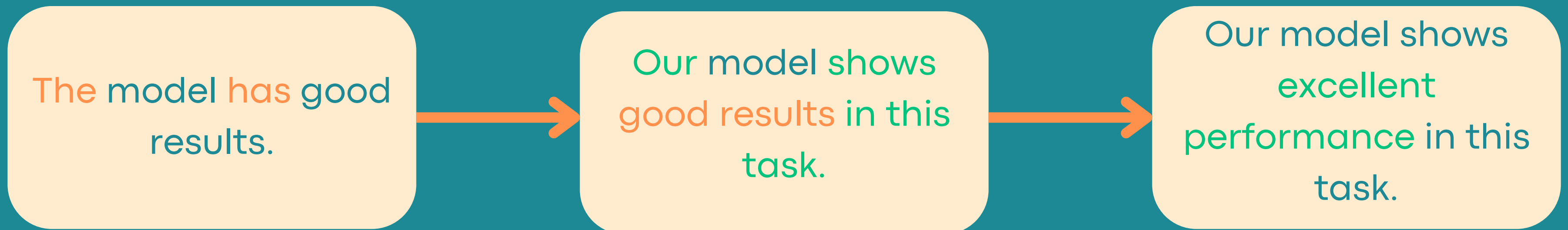
The paper will focus on the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy. The paper will also discuss the role of social media in the development and maintenance of mental health problems, and will discuss the implications of these findings for mental health practice and policy.

An exemple of introduction labeled with the CARS model

2 - The text revision task

Definition: Text revision is the transformation of an input text into an improved version fitting a desired attribute (formality, clarity, etc.), closer to the intended text

Example:



2 - The text revision task

3 types of tools:

01

Sentence
revision
tools

02

Grammar
checkers

03

Moves
annotators

3 - Currently available tools

Sentence revision tools

R3 →

Editing suggestions

CLARITYCOHERENCEFLUENCYMEANING-CHANGEDOTHERSSTYLE

current revision cycle: 1

Edit Intentions

65-86353-depth-1: Global crude oil prices spiked by over US\$4 per barrel yesterday after news came in ...

Global crude oil prices spiked by over US\$4 per barrel yesterday after news came in that a major pipeline BEFORE → FLUENCY pipeline AFTER in Minnesota had exploded and caught fire, killing two workers. The pipeline carries oil from Saskatchewan, Canada to close to Chicago, United States, and this oil represents 16\% of America's total oil imports. The section of BEFORE → FLUENCY the AFTER pipe that exploded was undergoing maintenance at the time to address a pinhole leak first identified three weeks earlier, and patched at the time with a metal sleeve. On Wednesday, the entire section was removed and replaced with a new one. However, when BEFORE → COHERENCE When AFTER oil was reintroduced on Thursday morning, oil leaked where the new section joined the pre-existing pipe, triggering the fatal explosion. The fire was extinguished BEFORE → FLUENCY extinguished AFTER later the same day. The line is split between four seperate BEFORE → FLUENCY separate AFTER pipes. After the explosion, all four BEFORE → FLUENCY pipes AFTER were shut down, resulting in the global price surge, but this subsided for the most part after it became clear that the three undamaged pipelines had returned to normal operation, restoring 80\% of the line's capacity, and that the accident pipe is expected to be back in use in a matter of days. Embridge identified the deceased as Dave Mussati Jr. and Arnovich, both of whom were contract workers based in Superior, Wisconsin. The damaged section is in Clearbrook, which is aproximatly BEFORE → FLUENCY approximately AFTER 350 miles west of Minneapolis. Embridge metallurgists have been sent to examine the failed section in an to determine the cause of the accident. Other Embridge workers are working to clear spilled oil from the site. It is expected that once this is complete, the pipe will return to service. Crude oil trading closed yesterday slightly above orriginal BEFORE → FLUENCY original AFTER prices

1/8

pieline BEFORE → FLUENCY pipeline AFTER

Replacement by FLUENCY

Prev

Reject or Accept

☐ Reject

☒ Accept

Confirm

Next

Submit

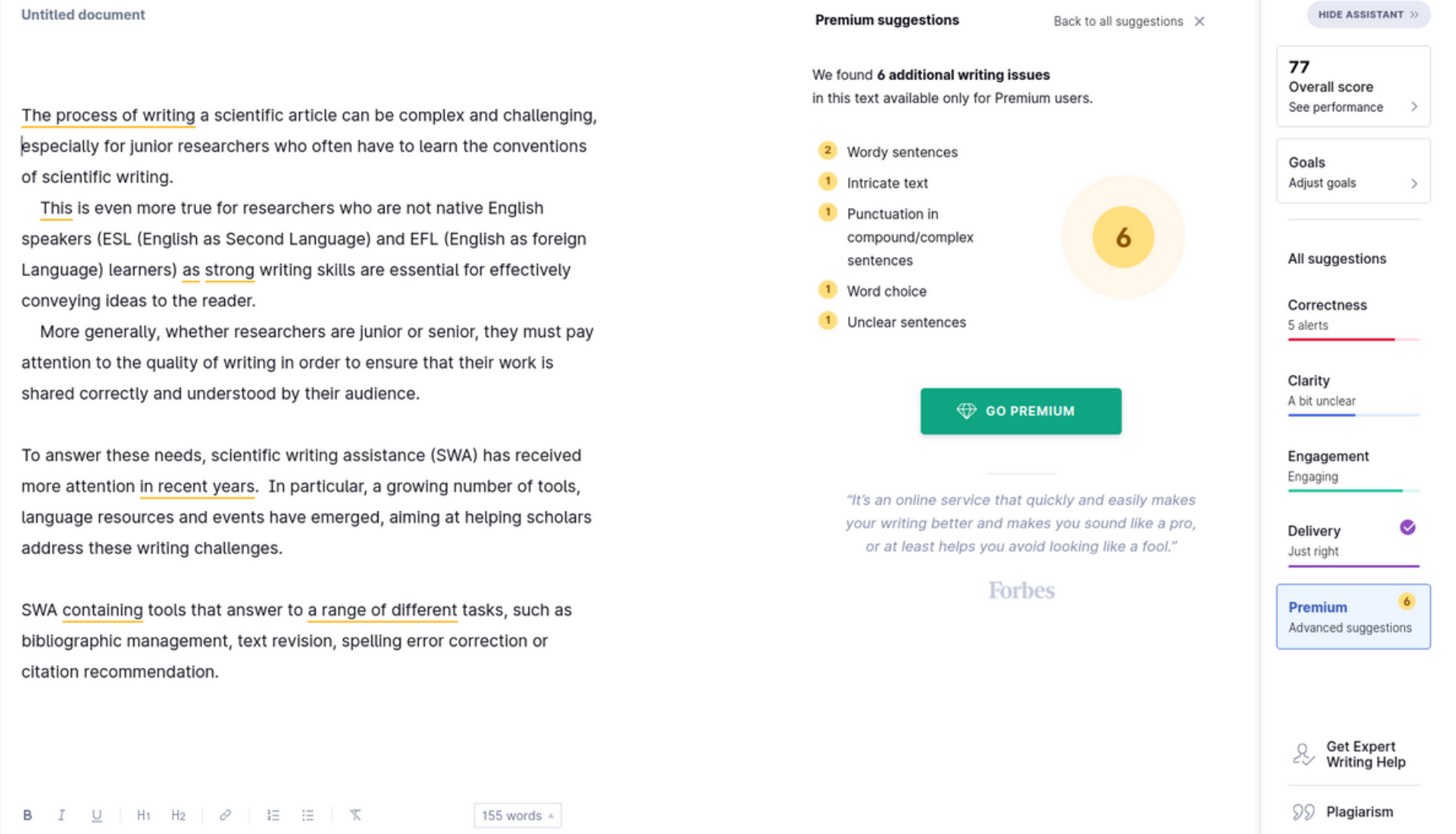
Interaction panel

Tool	Year	Domain	Approach	Availability
Langsmith	2020	Scientific	Transformer based	Free and paid plans
R3	2022	General/ Scientific		Open source
Chat GPT	2023	General		Free and paid plans

3 - Currently available tools

Grammar checkers

- grammar error correction (GEC)
- spelling error correction (SEC)



LinggleWrite ↑

Tool	Year	Domain	Approach	Availability
Grammarly	2023	General	Transformer based	Free and paid plans
LinggleWrite	2020	Academic	LSTM/ Bi-LSTM	Free to use

3 - Currently available tools

Move annotators

- dedicated to academic writing
- highlight the moves
- Help in the revision task by vizualisation



↑ RWT

Tool	Year	Domain	Approach	Availability
Mover	2016	Academic	Naive Bayes Classifier	Free to use
RWT	-	Academic	Probabilistic models	Limited access
AcaWriter	2022	Academic	Rule-based	Open source

4 – Future directions and challenges

- 1 Benchmark performance
 - Evaluation process to compare the tools' performances
- 2 Improve revisions models
 - Consider a larger context
 - Include the argumentative structure
- 3 Improve accessibility and transparency
- 4 Emergence of ethical issues

Text revision in Scientific Writing Assistance: An Overview

 Nantes
Université

LS2N



Léane Jourdan,
Florian Boudin,
Richard Dufour,
Nicolas Hernandez

Contact: leane.jourdan@ls2n.fr



ECIR
2023

BIR 2023

13th International Workshop on
Bibliometric-enhanced Information Retrieval

Bibliography 1/2

C. Strobl, E. Ailhaud, K. Benetos, A. Devitt, O. Kruse, A. Proske, C. Rapp, Digital support for academic writing: A review of technologies and pedagogies, *Computers & education* 131 (2019) 33–48.

E. D. Kallestinova, How to write your first research paper, *The Yale journal of biology and medicine* 84 (2011) 181.

S. Bourekkache, English for specific purposes: writing scientific research papers. Case study: PhD students in the computer science department, Master's thesis, University of Biskra, Algeria, 2022.

J. M. Swales, *Genre Analysis: English in academic and research settings*, The Cambridge applied linguistics series, The press syndicate of the University of Cambridge, 1990.

L. B. Sollaci, M. G. Pereira, The introduction, methods, results, and discussion (imrad) structure: a fifty-year survey, *Journal of the medical library association* 92 (2004) 364.

A. Seow, The writing process and process writing, *Methodology in language teaching: An anthology of current practice* 315 (2002) 320.

E. D. Laksmi, " scaffolding" students' writing in efl class: Implementing process approach, *TEFLIN Journal* 17 (2006) 144–156.

S. Bailey, *Academic writing: A handbook for international students*, Routledge, 2014.

R. A. Alves, T. Limpo, Progress in written language bursts, pauses, transcription, and written composition across schooling, *Scientific Studies of Reading* 19 (2015) 374–391.

E. A. Silveira, A. M. de Sousa Romeiro, M. Noll, Guide for scientific writing: how to avoid common mistakes in a scientific article, *Journal of Human Growth and Development* 32 (2022) 341–352.

Bibliography 2/2

W. Du, Z. M. Kim, V. Raheja, D. Kumar, D. Kang, Read, revise, repeat: A system demonstration for human-in-the-loop iterative text revision, in: Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 96–108.

J. Li, Z. Li, T. Ge, I. King, M. R. Lyu, Text revision by on-the-fly representation optimization, arXiv preprint arXiv:2204.07359 (2022).

T. Ito, T. Kuribayashi, H. Kobayashi, A. Brassard, M. Hagiwara, J. Suzuki, K. Inui, Diamonds in the rough: Generating fluent sentences from early-stage drafts for academic writing assistance, in: Proceedings of the 12th International Conference on Natural Language Generation, Association for Computational Linguistics, Tokyo, Japan, 2019, pp. 40–53.

L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, arXiv preprint arXiv:2203.02155 (2022).

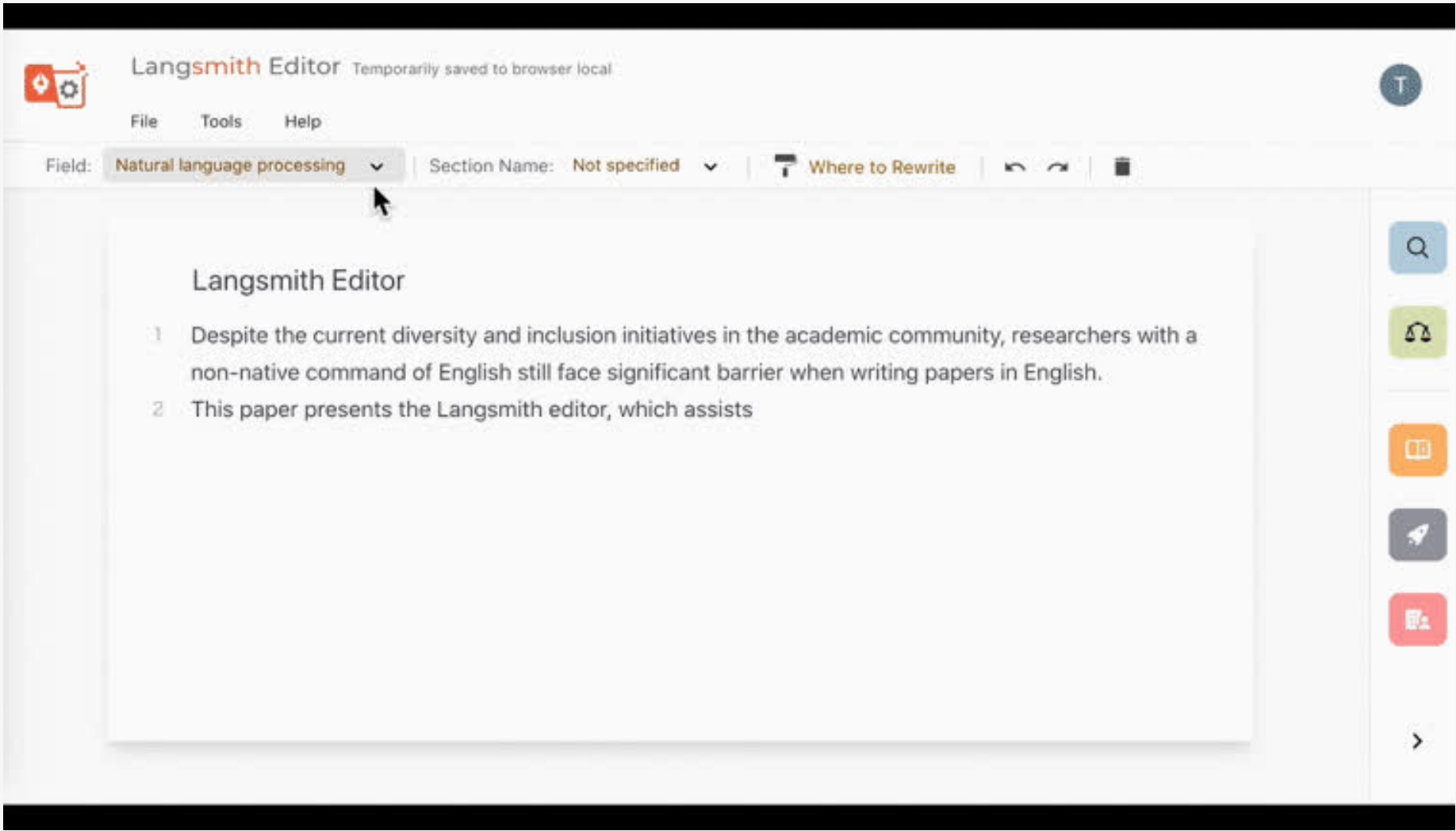
C.-T. Tsai, J.-J. Chen, C.-Y. Yang, J. S. Chang, LinggleWrite: a coaching system for essay writing, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 127–133.

L. Anthony, G. V. Lashkia, Mover: A machine learning tool to assist in the reading and writing of technical papers, IEEE transactions on professional communication 46 (2003) 185–193.

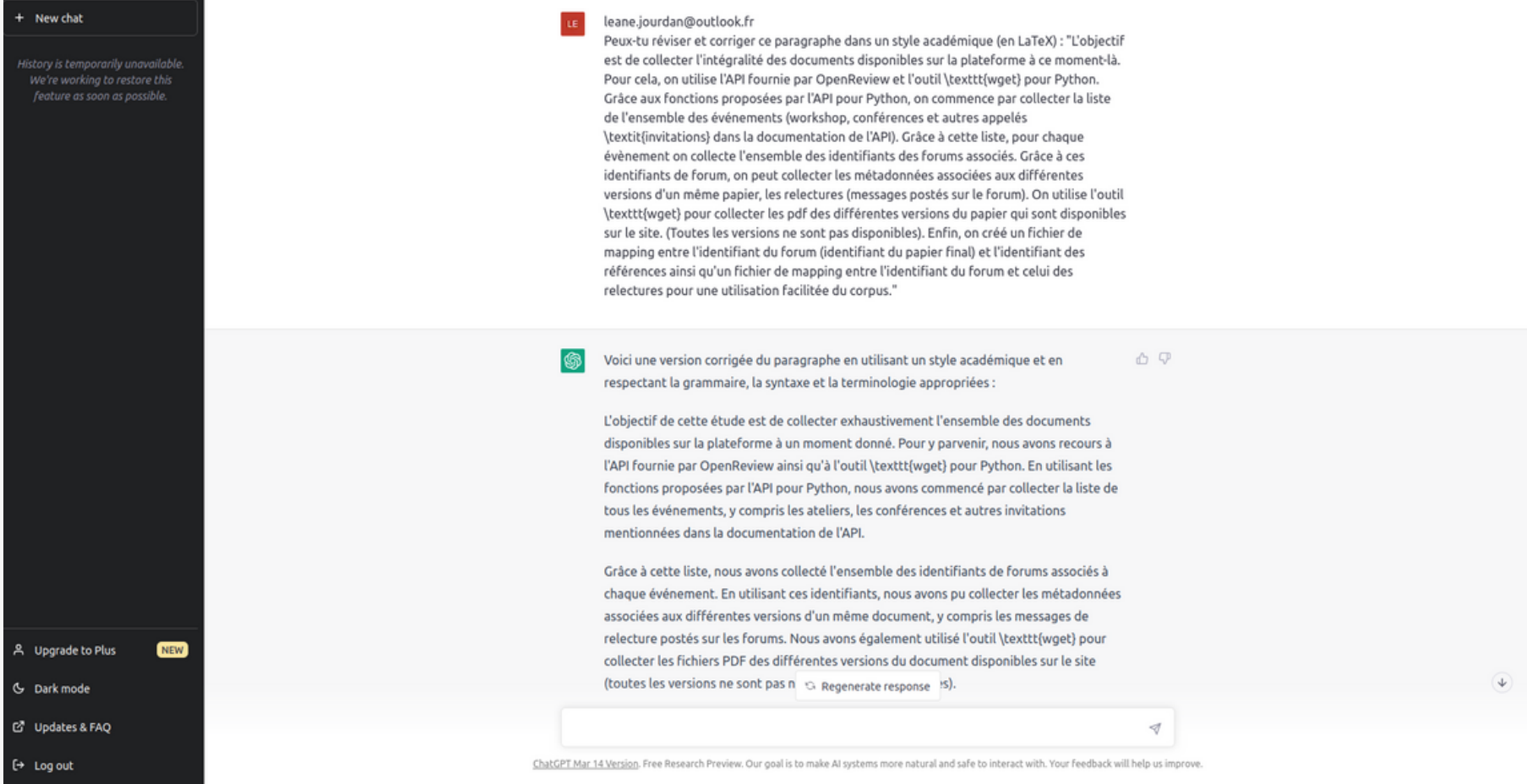
E. Cotos, Computer-assisted research writing in the disciplines, in: Adaptive educational technologies for literacy instruction, Routledge, 2016, pp. 225–242.

S. Knight, A. Shibani, S. Abel, A. Gibson, P. Ryan, Acawriter: A learning analytics tool for formative feedback on academic writing, Journal of Writing Research (2020).

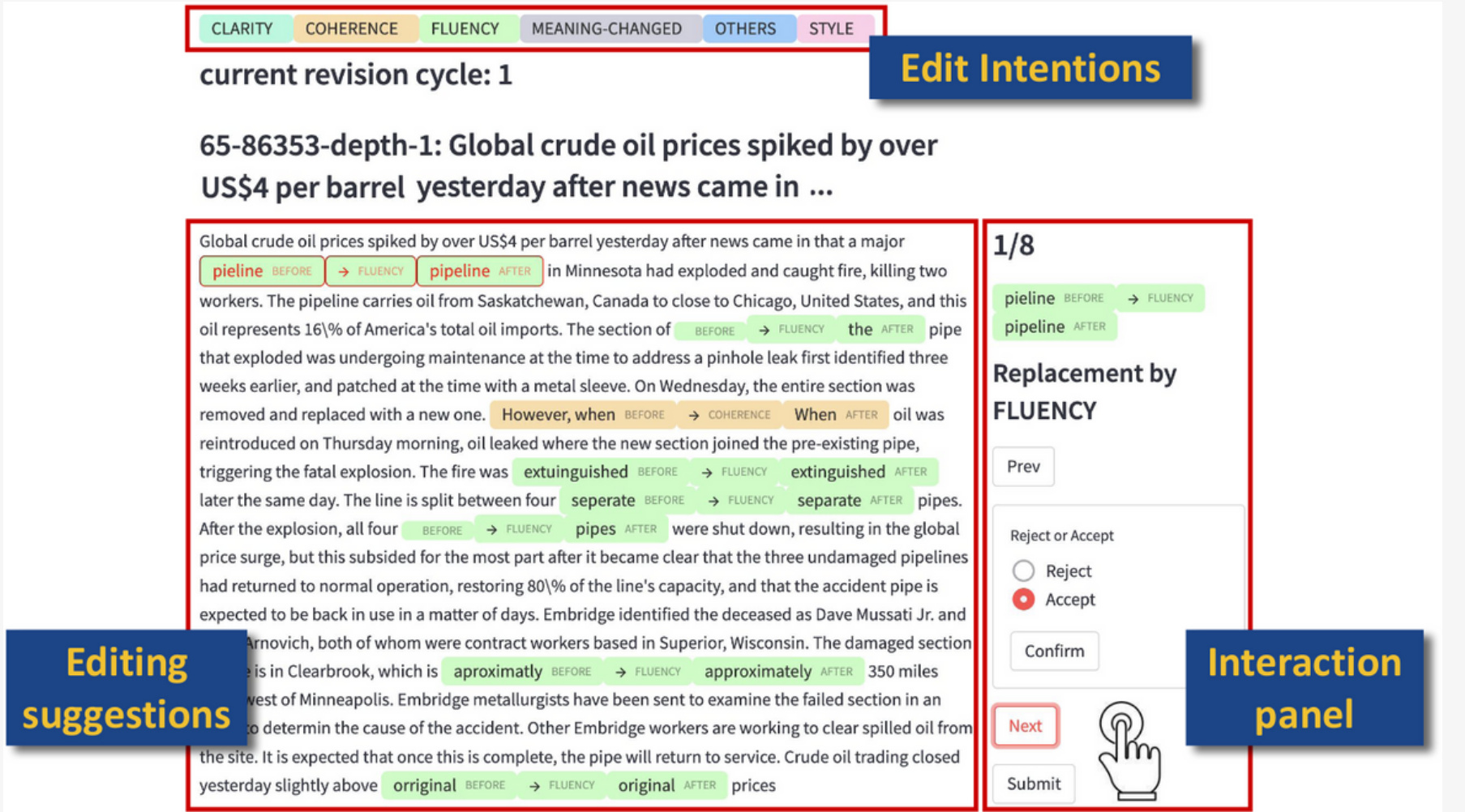
Sentence revision tools



Langsmith



ChatGPT



R3

Grammar checkers

Untitled document

The process of writing a scientific article can be complex and challenging, especially for junior researchers who often have to learn the conventions of scientific writing.

This is even more true for researchers who are not native English speakers (ESL (English as Second Language) and EFL (English as foreign Language) learners) as strong writing skills are essential for effectively conveying ideas to the reader.

More generally, whether researchers are junior or senior, they must pay attention to the quality of writing in order to ensure that their work is shared correctly and understood by their audience.

Premium suggestions

Back to all suggestions

We found 6 additional writing issues in this text available only for Premium users.

2

Wordy sentences

1

Intricate text

1

Punctuation in compound/complex sentences

1

Word choice

1

Unclear sentences

GO PREMIUM

“It’s an online service that quickly and easily makes your writing better and makes you sound like a pro, or at least helps you avoid looking like a fool.”

Forbes

77

Overall score

See performance

Goals

Adjust goals

All suggestions

Correctness

5 alerts

Clarity

A bit unclear

Engagement

Engaging

Delivery

Just right

Premium

Advanced suggestions

Get Expert Writing Help

Plagiarism

B I U H1 H2

155 words

Grammarly

I received your letter, so I am writing to you to give you some informations about me. I would like to travel on July because I finish school on June and I go to the Great Britain August for one month.

Sentence-level feedback

This seems to be a good sentence.

This sentence could maybe be improved.

There are some problems in this sentence.

Check again

Keep writing

Writing Suggestion

FINISH

[V n] 32

V school, V job, V sentence

She finished (the concert) with a song from her first album.

她以自己第一張專輯中的一首歌作為（音樂會的）結尾。

[V adv] 15

V third, V first, V fourth

She finished second (= in second place) in the finals.

她在決賽中獲得了第二名。

[V -ing] 8

V reading, V eating, V talking

Have you finished reading that magazine?

你看完那本雜誌了嗎？

[V with n] 7

V with point, V with yard, V with record

The play finishes with a wedding.

這齣戲以一場婚禮結束。

Writing Proficiency

Help

CEFR Level

A2

A1 A2 B1 B2 C1 C2

Select a sentence to detect grammatical errors

C

I received your letter, so I am writing to you to give you some informations about me. I would like to travel on July because I finish school on June and I go to the Great Britain August for one month.

Click an error marked with

Insert

Delete

Replace

 to receive suggestions

D

I would like to travel on July because I finish school on June and I go to the Great Britain INSERT August for one month.

school_ June

N-gram

Percent

Count

Example

school in June

62.1 %

7,933

Show

From the date upon which I am writing until the close of school in June the work will consist of longer runs as the speed of the students can safely be increased .

In a major speech at the Central Party School in June last year , Hu exhorted

LinggleWrite

AntMover 1.0 (Structural Analysis Software) Laurence Anthony, 2003

File Settings Tools Project: computer_science_abstracts

Directory antmover
antmover_testing_data\antmover_demo.txt

Class 2 (Making_Topic_Generalizations) 2nd Opinion
Teaching learners about the common structural patterns used in different types of texts , such as the abstract and introduction of research papers , has proved successful in many ESP reading and writing courses .

Class 5 (Indicating_a_Gap) 2nd Opinion
However , a major problem faced by researchers when analyzing texts is the vast amount of time needed to conduct the analysis .

Class 2 (Making_Topic_Generalizations) 2nd Opinion
This has led to many studies reporting only preliminary findings , based on a small corpus of target texts .

Class 9 (Announcing_Present_Research) 2nd Opinion
In this paper , we propose a computer system that uses machine learning to automatically identify the structure of texts , enabling researchers to quickly and effectively process very large corpora .

Class 11 (Evaluation_of_Research) 2nd Opinion
The system also has applications in the classroom as a teacher resource when evaluating and selecting texts that highlight certain features , and as a student resource when conducting data - driven learning .

Log Finished creating feature probabilities
Active project is computer_science_abstracts

Original Processed Moves Outline Optimize flow Add to Training Delete Delete All Exit

Mover

AcaWriter

Introduction	Methods	Editing Results	Discussion
--------------	---------	------------------------	------------

Enter or edit text below:

Click text below for feedback

Analysis

EXPERIMENTAL STUDY OF THE WEB-BASED CALL

Twenty university students (non-native English speakers, majoring in Computer Science) from Southwest Normal University in China participated in this experiment via the Internet. The participants were randomly divided into two groups and assigned to different treatment conditions (see Table 3). The first group was assigned to the intra-personal treatment (i.e., where language is used for a functional purpose) comprised of a note-taking task and a dictation task (abbreviated as N + D). The other group was assigned to inter-personal treatment (i.e., where language is used for a communication purpose) comprised of an information-gap task (abbreviated as IG). **Materials**

The prototype course consisted of 20 dialogues. Each dialogue contained three videos from simple to complex (specified to support the same instructional goal, though by different dialogue strategies)¹ and a set of questions in different task types as described in the Course Structure section, above. The structured or unstructured input video clips are based on the topic "Search"; The task types are defined as dictation, note-taking, and information-gap. The lesson contents were selected based on the participants' current knowledge level and course strategies. The selection of level of the video for each dialogue was left to the students themselves. The default level was based on the participants' progress. However, if the participants were not satisfied with the default level, they were able to select their own preferred level (see video levels in Figure 2).

Words: 1028 Goal: 636 to 2385

Change Draft: Draft2 2013/10/09 14:01

ANALYZE

Export Options

Instrumentation

The independent variable was defined as type of treatment. The dependent variables were achievement and attitude. The achievement variable was used to address the extent of learning associated with the two treatments, N + G and IG, respectively. The attitude variable was employed to address the degree of motivation reported for the Web-based communication platform as well as the different tutorial functions.

The tracked data (i.e., the participants' written responses to each question) were used to measure the participants' levels of achievement. The first session was used to acquaint the participants with the training process and was therefore not scored. Three of the remaining 19 dialogues were randomly selected to be scored and were used to measure the participants' achievement in L2 learning in the Web-based CALL environment.

Finn's (1977) type and token analysis was employed to score participant responses. Hunt's (1977) T-unit2 word count was also used to measure their syntactic complexity. Types are the number of different words used in the participants' responses. The tokens are the total number of words written. The number of types reflects a direct measure of the breadth of subject-specific vocabulary items acquired. The token analysis was conducted on T-units. The assumption is that the length of the T-unit increases as learners mature or develop intellectually (Hunt, 1977).

FEEDBACK (Click on text above for sentence level feedback)

You are likely providing general orientation, giving theoretical, empirical, or informational background relevant to the results of your study here.

Comments

MOVE 1. APPROACHING THE NICHE

You: 93%

0% - 4% 4% - 11% 11% - 14%

not enough goal too much

1 step(s) needs work | 2 step(s) good work

MOVE 2. OCCUPYING THE NICHE

You: 6%

26% - 43% 43% - 77% 77% - 94%

not enough goal too much

1 step(s) needs work | 1 step(s) good work

MOVE 3. CONSTRUING THE NICHE

You: 1%

0% - 6% 6% - 22% 22% - 30%

not enough goal too much

4 step(s) needs work | 1 step(s) good work

RWT