

Tutoriel : relire des articles en TAL

Kevin Bretonnel Cohen, **Karën Fort**, Margot Mieskes, Aurélie Névéal,
Anna Rogers

TALN – 29 juin 2021

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

La relecture

Qu'est-ce qu'une bonne mauvaise (plus facile) relecture ?

Du point de vue des auteurs :

- lecture superficielle du papier
- impolitesse ou méchanceté
- inflexibilité
- attendre la perfection
- relire votre version du papier plutôt que celle soumise : intérêt dans le sujet, ressenti sur le papier, ...
- faire des remarques infondées sur la papier
- [insérer vos mauvaises expériences ici]

Sources : <https://acl2017.wordpress.com/2017/02/23/last-minute-reviewing-advice/> et <https://sites.umiacs.umd.edu/elm/2016/02/01/mistakes-reviewers-make/>

Qu'est-ce qu'une bonne mauvaise (plus facile) relecture ?

Du point de vue des responsables de *track* (*chairs*) :

- une relecture **en retard** (sans prévenir et planifier)
- une relecture **moyenne**
- une relecture positive avec des scores faibles (une relecture **incohérente**)
- une relecture **insultante**
- [insérer vos mauvaises expériences ici]

Sources : <https://acl2017.wordpress.com/2017/02/23/last-minute-reviewing-advice/> et <https://coling2018.org/reviewer-code-of-conduct/>

Exercice : faire une relecture d'article



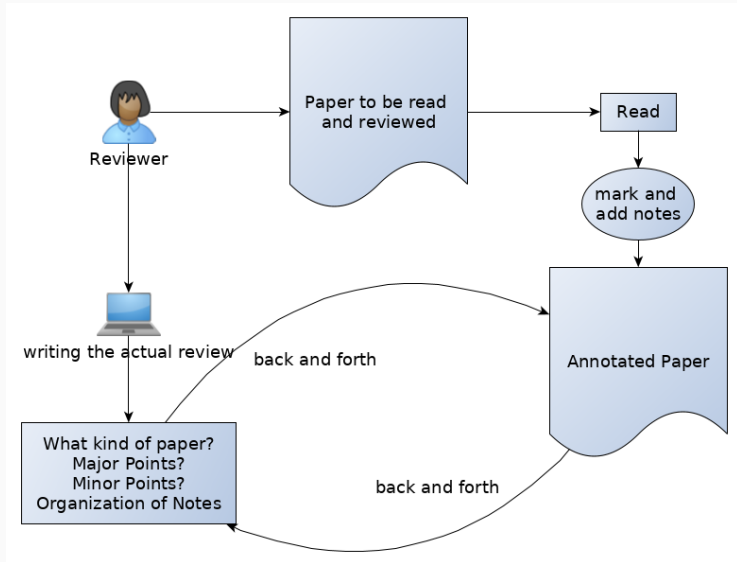
- Lisez le **Papier 42**
- Ecrivez une courte relecture (notes)

Ce que vous devez vous demander

- Est-ce que j'ai pris assez de temps pour digérer le papier ?
- Est-ce que j'ai pris assez de temps pour écrire ma relecture ?
- Est-ce que je me suis bien concentré sur le travail ?
- Est-ce que mes commentaires sont constructifs ?
- Est-ce que mes commentaires sont assez substantiels (sont-ils justifiés) ?
- Est-ce que j'ai mentionné quelque chose de bien sur le papier ?
- Est-ce que j'ai vérifié :
 - ▶ les données et le code fournis (éventuellement) ?
 - ▶ le plagiat (l'auto-plagiat) ? le saucissonnage (*salami-slicing*) ?

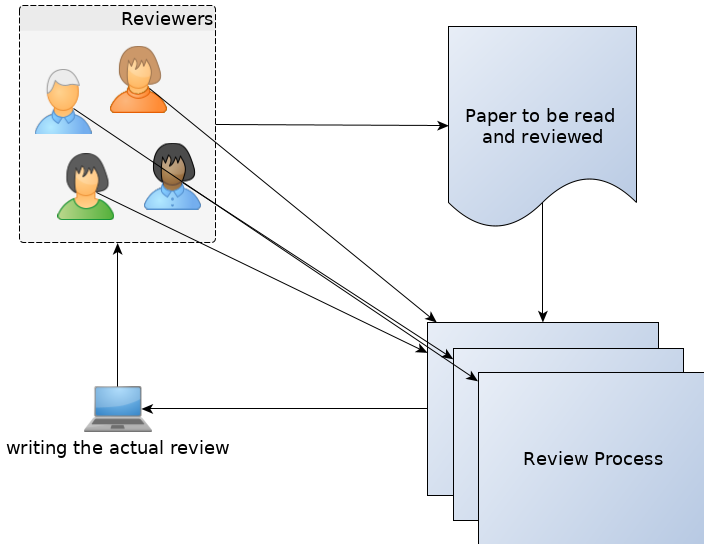
?? Est-ce que c'est une relecture que j'aimerais recevoir ?

Comment avez-vous procédé ? une possibilité...





Beaucoup d'autre possibilités





Paper to Review

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptus. **At vero eos et accusam et justo duo dolores et ea rebum.** Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptus. **At vero eos et accusam et justo duo dolores et ea rebum.** Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptus. **At vero eos et accusam et justo duo dolores et ea rebum.** Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Main point

Strength

Weakness

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et justo odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugiat nulla facilisi. **Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptus.**

Weakness

Strength

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit laboris nisi ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in ~~vulputate velit esse~~ molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et justo odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugiat nulla facilisi.

tytu

grammar

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer



- De quel type de recherche s'agit-il ?
 - ▶ sur la méthodologie
 - ▶ sur l'évaluation
 - ▶ sur la reproductibilité
 - ▶ sur des ressources langagières
 - ▶ papier de positionnement ou d'état de l'art



Identifier le type de papier et utiliser les critères d'évaluation appropriés



- Can you tell what research question was addressed?
 - ▶ What is the hypothesis?
 - ▶ How was it tested?
- Is the research technically sound and correct?
- Is there an error analysis?
- Is there meaningful comparison?
 - ▶ To the literature?
 - ▶ To baselines?



Reproduction paper

- Data/Code availability
 - ▶ Of the original research
 - ▶ Of the reproduction
- Description of the specific reproduction obstacles
- Generalizability : beyond reproduction ?
- Informativeness : insight of the reproduction ?
- Is there meaningful comparison to the original experiments ?



Resource paper

- How was the resource produced ?
 - ▶ Source material ?
 - ▶ Workforce ?
- How was the resource evaluated ?
 - ▶ What was the quality control procedure ?
 - ▶ Was inter-annotator agreement computed ?
- Are metadata and accessibility detailed ?
- How does it compare to existing resources ?



Survey paper

- Organization : does the survey address a specific question ?
- Scope : Does the paper identify a reasonably focused area to survey ?
- Thoroughness : does the paper cover all of the relevant literature ?
 - ▶ Is the source of the literature surveyed clear ?
 - ▶ Are criteria for inclusion/exclusion of papers in the survey clear ?
- Outlook : Does the paper identify areas for future work ?



Plus généralement

- De quoi parle le papier ?
- Qu'est-ce qui est utilisé ?
 - ▶ Données
 - ▶ Méthodes
 - ▶ Outils
 - ▶ ...
- Quels sont les résultats ?
- Quelles sont les revendications des auteurs ?
- Est-ce que les résultats et les revendications correspondent ?



Première (ou dernière ?) étape : écrire un résumé !



Plus généralement

- Est-ce que les hypothèses sont solides ?
- Est-ce que le protocole expérimental est adapté au problème traité ?
- Est-ce que l'évaluation est solide ?
- Est-ce que le problème est correctement décrit ?
- Quel est le message à retenir ?
- Est-ce que les conclusions correspondent aux résultats obtenus ?
- ...



- Suivre l'ordre du papier
- Lister les forces (et les faiblesses ?)
- Problèmes importants vs. mineurs

Vos affirmations doivent être **étayées**

Les formulaires de relecture

La plupart des conférences du domaine utilise le même genre de formulaire contenant :

- des scores numériques
- des champs de texte libre (avec, ou pas, un minimum)
- des commentaires confidentiels (fac.)

Exercice : ce fragment de relecture pose problème, pourquoi ?



- **Score** : Etat de l'art : 1/5 (des références majeures manquent)
- **Commentaire aux auteurs** : il manque des références / l'état de l'art est incomplet.

Exercice : ce fragment de relecture pose problème, pourquoi ?



- **Score** : Etat de l'art : 1/5 (des références majeures manquent)
- **Commentaire aux auteurs** : il manque des références / l'état de l'art est incomplet.

Comment l'améliorer ?



Scores (attention !)

- Doivent être **cohérents** avec vos commentaires
- Toujours vérifier



Remplir le formulaire en ligne vs hors ligne

En ligne

- + facile à remplir
- temps limité (vous risquez de perdre vos scores/vos commentaires)

Hors ligne

- + pas de limite de temps
- peu se révéler pénible (vérifier l'import)

Discussion clipart : MRafizeldi / CC BY-SA



Remplir le formulaire en ligne vs hors ligne

En ligne

- + facile à remplir
- temps limité (vous risquez de perdre vos scores/vos commentaires)
- biais dû au formulaire ?

Hors ligne

- + pas de limite de temps
- peu se révéler pénible (vérifier l'import)
- biais dû au formulaire ?

→ Approche mixte (?) : écrivez vos commentaires dans un fichier texte, puis copiez/collez les dans le formulaire



Les commentaires confidentiels



Source : Free photo by Tophee Marquez from <https://www.pexels.com>



Les commentaires confidentiels

- Circonstances particulières qui pourraient justifier d'accepter/rejeter le papier malgré des scores faibles/élevés
- Commentaires sur les auteurs plutôt que sur le papier
- Doutes concernant du plagiat / saucissonnage
- Doutes concernant votre relecture (niveau d'expertise, conflit d'intérêt...)
- Brèches dans l'anonymat (y compris *le vôtre* : peut-être avez-vous déjà relu ce papier...)

N'importe quoi d'autre qu'il **vous** semble important de faire remonter et qui ne doit **pas** être visible aux autres relecteurs et/ou aux auteurs



(très) mauvais papiers ?



Ne pas **perdre son temps**

vs

faire un **effort pédagogique**

Discussion clipart : MRafizeldi / CC BY-SA



(très) mauvais papiers ?



Ne pas **perdre son temps**

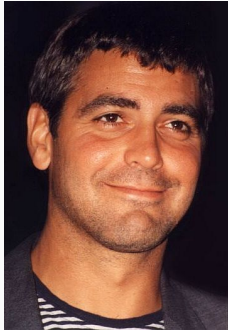
vs

faire un **effort pédagogique**

→ avez-vous eu l'impression que les auteurs ont fait des efforts ?

Ne pas oublier

What else ?



CC-BY-SA-2.0 / John Mathew Smith

Plagiat

« **appropriation** des données d'autrui, de résultats de recherche ou de mots sans leur octroyer le crédit approprié » (Rapport [Alix](#)).

Exercice : retour au Papier 42



- Relisez le Papier 42
- Est-ce que, par hasard, il ne contiendrait pas du plagiat ?
- Si vous pensez que c'est le cas, trouvez la source

Des logiciels de **détection de plagiat** sont souvent disponibles dans votre université, par ex. Compilatio (Sorbonne Université), iThenticate (University of Amsterdam), etc. . .

Ou utilisez votre propre système de **détection de similarité**

Ou utilisez votre **moteur de recherche** préféré

...à la détection du saucissonnage (*salami-slicing*)

Un effet de bord de chercher d'éventuels plagiats (rares) est que vous risquez de tomber sur du saucissonnage (ou de l'auto-plagiat) :

"Salami publication (sometimes called bologna or trivial publication) is the practice of dividing one significant piece of research into a number of small experiments [...] to increase the number of publications."

<https://www.law.hku.hk/researchintegrity/salami-slicing/>

Exercice : matériel fourni avec le Papier 42



- est-ce que du matériel additionnel est fourni ?
- est-ce qu'il est pertinent ?

Lorsque des données (du code) sont fournies :

- vérifiez qu'elles existent
- vérifiez qu'elles ne cassent pas l'anonymat (si c'est le cas, le signaler aux *chairs*)
- si vous avez le temps (jamais), essayez de télécharger et de faire tourner



Ecrire une relecture

- Prévoir assez de **temps**
- Rester **poli**
- Ne considérer que le **travail**
- Eviter les trivialités
- Ne pas attaquer les auteurs
- Prendre en compte les particularités (par ex. langue peu dotée)
- Etre aussi **précis** et **explicite** que possible
- Plus les scores donnés sont bas, plus la relecture doit être détaillée pour les justifier

Merci de votre participation !