Réécritures d'un mythe et outils de détection des réutilisations. De l'Orphée de Virgile à celui de Ballanche

Mythological Rewriting and Textual Similarity Detection Tools: from Virgil's Orpheus to Ballanche's

Karolina Suchecka et Nathalie Gasiglia

Édition électronique

URL : <https://journals.openedition.org/revuehn/2467>

DOI : 10.4000/revuehn.2467

ISSN : 2736-2337

Éditeur

Humanistica

Référence électronique

Karolina Suchecka et Nathalie Gasiglia, « Réécritures d'un mythe et outils de détection des réutilisations. De l'Orphée de Virgile à celui de Ballanche », *Humanités numériques* [En ligne], 4 | 2021, mis en ligne le 01 décembre 2021, consulté le 02 décembre 2022. URL : <http://journals.openedition.org/revuehn/2467> ; DOI : <https://doi.org/10.4000/revuehn.2467>

Creative Commons - Attribution 4.0 International - CC BY 4.0
<https://creativecommons.org/licenses/by/4.0/>

Réécritures d'un mythe et outils de détection des réutilisations. De l'Orphée de Virgile à celui de Ballanche

*Mythological Rewriting and Textual Similarity
Detection Tools: from Virgil's Orpheus to Ballanche's*

Karolina Suchecka et Nathalie Gasiglia

Résumés

Un corpus de réécritures du mythe d'Orphée et Eurydice nous permet d'étudier l'intertextualité et de la représenter au sein d'une édition numérique comparative par le biais d'un réseau de correspondances. Pour ce faire, nous exploitons deux logiciels de détection des réemplois textuels, TextPAIR et Tracer, pour combiner leurs traitements et exploiter au mieux les possibilités de chacun. Cela ne va pas sans poser quelques difficultés, que nous exposons avant de proposer une série de traitements et d'enrichissements. Ces manipulations techniques nous permettent d'interpréter les résultats obtenus pour un essai de Pierre-Simon Ballanche, en les confrontant à des analyses empiriques du topos orphique. Ainsi, nous montrons que l'analyse outillée est d'autant plus utile pour l'étude de l'intertextualité qu'elle est adaptée au corpus traité.

A corpus collecting rewritings of the myth of Orpheus and Eurydice allows us to study intertextuality and to represent it through a network of correspondences and within a comparative digital edition. For this purpose, we examine the output of two text reuse detection tools, TextPAIR and Tracer, in order to combine their treatments and exploit each one's best potential. We propose a series of treatments to enrich the results and to overcome specific challenges observed. These technical manipulations allow us to interpret the results obtained for an essay by Pierre-Simon Ballanche and confront them with empirical analyses of

the Orphic topos. Thus, we show that analysis supported by computational techniques is all the more useful for the study of intertextuality as it is adapted to the corpus.

Entrées d'index

MOTS-CLÉS : intertextualité, littérature, réécriture, réutilisation textuelle, TEI, critique littéraire, traitement automatique des langues

KEYWORDS: intertextuality, literature, rewriting, text reuse, TEI, literary criticism, natural language processing

- ¹ Disposer de corpus numériques de textes littéraires accélère le repérage de réemplois d'extraits, de tournures, de références, etc. et en améliore l'exhaustivité¹. Notre corpus, hétérogène et intermédial, est constitué de réécritures du mythe d'Orphée et Eurydice. Afin d'en faire une édition numérique comparative enrichie d'un réseau des correspondances détectées entre les œuvres, nous usons de méthodes et d'outils littéraires, informatiques et linguistiques.
- ² Les logiciels de détection de réemplois textuels, comme TextPAIR² et Tracer³, exploitent des techniques d'analyse de séquences qui ont divers usages, du séquençage du génome à la détection du plagiat⁴. Basés sur des techniques de l'intelligence artificielle, comme la vectorisation, ils repèrent bien mieux les correspondances proches que les réutilisations et évocations complexes.
- ³ Or, notre corpus regroupe plus de 70 réécritures françaises⁵ de genres et d'époques différents (en ancien français et en français moderne), dont la proximité au mythe est très variable. Nous les traitons avec les deux logiciels déjà évoqués, afin d'exploiter au mieux les possibilités de chacun, mais cela pose quelques difficultés. Nous les pallions au moyen de post-traitements appuyés sur des enrichissements. Enfin, nous exploitons les résultats obtenus pour explorer les réseaux des correspondances établies.

Présentation du corpus

- ⁴ Parmi une trentaine de mentions du mythe recensées dans les littératures grecque et romaine antiques (Béague *et al.* 2019), *Les Géorgiques* de Virgile (37-30 av. J.-C.) et *Les Métamorphoses* d'Ovide (I^{er} s. apr. J.-C.), dont la trame du mythe est commune, sont repris le plus fréquemment dans les réécritures. Orphée y est présenté principalement comme l'idéal de la fidélité conjugale et un chantre-magicien, alors que d'autres sources valorisent sa fonction de législateur et de fondateur des cultes.

Textes du corpus

- ⁵ Notre corpus inclut des traductions françaises des livres X et XI des *Métamorphoses* et du livre IV des *Géorgiques* (mais pas d'autres sources anciennes, souvent très fragmentaires) et des réécritures d'inspirations très diverses (non limitées aux textes antiques), qui, par leur hétérogénéité, offrent des défis pour la détection de leurs correspondances.
- ⁶ Tout d'abord, plusieurs « chaînes de réécriture » peuvent être observées, comme pour l'opéra *Orphée et Eurydice* (Gluck, Calzabigi et Moline 1774), qui s'inspire des sources anciennes et, sans doute et entre autres, d'*Orfeo* (Monteverdi et Striggio 1609), et qui a été à son tour parodié dans l'opéra-comique *Le Petit Orphée* (Deshayes et Rouhier-Deschamps 1792).
- ⁷ Ensuite, le rapport au mythe dépend du type de réécriture : alors que les parodies ridiculisent la musique d'Orphée ou son amour pour Eurydice, les modernisations incluent des amalgames, des personnages additionnels ou des décors modernes qui réorientent l'interprétation du mythe (cf. notamment Brunel 1997⁶).
- ⁸ Enfin, chez certains auteurs, Orphée est une figure récurrente, le mythe se modulant et devenant une constante créatrice, parfois implicite, parfois centrale. Pierre-Simon Ballanche, notamment, se réfère au mythe orphique pour prôner la force civilisatrice de la musique (1801), pour déplorer l'espoir perdu de la fidélité conjugale (1809) et pour conter les exploits d'un Orphée Argonaute (1819). Enfin, ces motifs, réunis, étoffés et enrichis, illustrent la doctrine sociale palingénésique de l'auteur dans l'épopée *Orphée* (1827). Le chantre y est prophète de liberté et d'égalité, successeur de Prométhée et fondateur des cultes.
- ⁹ Nous nous appuyons ici principalement sur un texte de cet auteur, *Sixième Fragment* (1809 [BALLANCHE1809 dans le corpus]). Ce bref essai se réfère au mythe, sans le parodier ni le moderniser. Malgré des reformulations et une poétisation assez importantes, nous montrons que plusieurs relations sont détectables avec des traductions des deux sources, surtout celle de Virgile. Nous dégageons aussi que, 20 ans plus tard, Ballanche réécrit des extraits de cet essai dans sa longue épopée *Orphée* (1827-1829 [BALLANCHE1829]), ce qui l'éloigne du texte de Virgile, et va jusqu'à modifier les emblèmes majeurs du mythe, comme la descente aux Enfers et la mort d'Eurydice⁷. Alors que, dans *Sixième Fragment*, Eurydice symbolise, comme dans le mythe, la douceur du mariage et l'inspiration poétique, dans *Orphée*, elle devient beaucoup plus active, initiatrice et elle-même créatrice. L'observation de la place de BALLANCHE1829 au sein du réseau des correspondances détectées nous permet aussi d'isoler et de relier des œuvres qui s'inscrivent dans le même esprit, notamment en valorisant un Orphée fondateur des cultes.
- ¹⁰ Avant de voir si ces relations complexes et les particularités de traitement du mythe peuvent être révélées par les logiciels de détection des réemplois textuels Tracer et TextPAIR, nous les présentons rapidement.

Détection des correspondances avec TextPAIR et Tracer

¹¹ TextPAIR détecte des correspondances selon des paramètres modulables (choix du nombre minimal des entités communes à détecter et de leur nature – mots pleins, lemmes ou stemmes (mots racinisés) –, liste des mots à ne pas prendre en compte...) et traite des fichiers XML, mais sans en exploiter le balisage. Dans un premier temps, l'outil génère des *n-grammes*⁸ qui se chevauchent pour chaque texte, les stocke et les indexe afin de les confronter à ceux des autres textes.

¹² Tracer, lui, traite des textes en format TXT, prédécoupés en segments (des phrases pour notre corpus) dotés d'un identifiant unique⁹. Des paramétrages sont possibles (recherche sur les lemmes ou uniformisation des mots similaires...) et le poids de chaque correspondance est calculé.

¹³ TextPAIR et Tracer semblent être très différents, tout en proposant des paramétrages complémentaires. Comme nous tâchons de le démontrer dans la suite de notre article, leurs résultats peuvent être combinés utilement, grâce, notamment, aux différences du traitement des textes soumis. Pour les deux logiciels, la documentation est riche et détaillée, les développeurs actifs et à l'écoute des utilisateurs, et le code source librement disponible. Enfin, des recherches littéraires très diversifiées ont été menées avec les deux outils, sur des corpus de langues, d'époques et d'auteurs différents¹⁰. Tous ces facteurs nous amènent à focaliser notre recherche sur ces deux outils en particulier, sans que cela nous empêche d'apprécier les méthodes et techniques d'autres outils¹¹ que nous n'exploitons pas directement, mais auxquels nous nous référons notamment en retravaillant les résultats initiaux de Tracer et TextPAIR.

¹⁴ Quel que soit le choix des outils, la comparaison et la compilation des sorties des deux logiciels ne sont pas toujours aisées. C'est pour cela que nous adoptons d'abord des paramètres d'analyse proches : recherche par trigrammes, lemmatisation, aplatissage des accents, taille minimale des correspondances de 3 mots et ordre des mots non pris en compte. Après avoir pallié quelques problèmes liés à l'exploitation combinée des logiciels, nous différencions progressivement leurs paramétrages. Ci-après, nous nous concentrons sur trois difficultés relatives à la préparation du corpus, la pertinence des résultats et le format des sorties. Nous n'abordons pas les prétraitements (établissement des lexiques, séquençage des phrases, etc.), qui jouent cependant sur la quantité et la qualité des résultats.

Traitement avec les deux logiciels

¹⁵ Notre corpus étant destiné à produire une édition numérique savante, enrichie des correspondances intertextuelles (repérées avec TextPAIR et Tracer), il est structuré en conformité avec le standard XML-TEI P5, qui permet un balisage adapté à chaque type d'œuvres : pièces de théâtre et livrets d'opéra, poésie et prose¹².

Prétraitement du corpus et pertinence des résultats

16 Comme nous l'avons vu ci-dessus, aucun des logiciels ne sait exploiter le balisage du corpus, alors que cela aurait amélioré les résultats¹³. En outre, la segmentation des textes en phrases limite la recherche des correspondances de Tracer, alors que TextPAIR est, *a priori*, capable de détecter des segments communs plus étendus et des reformulations incluses dans de longs fragments. C'est avéré pour des réemplois très proches, mais pour des correspondances plus subtiles, au contraire, TextPAIR trouve des extraits très courts, reliés par des syntagmes ou des constructions polylexicales proches. Tracer est ainsi, finalement, plus apte à détecter des reformulations fines au sein de phrases longues, que TextPAIR omet parce que les entités communes sont trop éloignées l'une de l'autre¹⁴. En contrepartie, Tracer détecte beaucoup de relations peu pertinentes entre des phrases courtes (« Orphée reste seul » et « Junon reste seule avec Jupiter »), que TextPAIR ne retient que si au moins une autre entité commune est trouvée dans le contexte proche.

Formats des sorties

17 Le choix de faire coopérer deux logiciels indépendants nécessite un post-traitement afin que leurs sorties soient compilables et comparables. Pour BALLANCHE1809, par exemple, TextPAIR repère 20 correspondances et Tracer en repère 40. Ils détectent la plupart du temps les mêmes couples, mais leurs présentations diffèrent : une liste de propriétés associées à une valeur pour TextPAIR (figure 1) et un fichier tabulé pour Tracer (tableau 1), formats mieux lisibles par une machine que par un humain.

FIGURE 1. EXTRAIT D'UNE CORRESPONDANCE DÉTECTÉE PAR TEXTPAIR

```
{
  "source_passage": "touchaient au seuil du séjour des vivants. Mais,
    ô faiblesse d'un cœur qui aime ! Orphée s'arrête pour écouter",
  "source_author": "Pierre-Simon de Ballanche",
  "source_title": "Sixième Fragment",
  [...]
  "target_passage": "touchait au seuil qui sépare l'empire des mots du
    séjour des vivants, lorsqu'il s'arrêta involontairement",
  "target_author": "Pierre-Simon de Ballanche",
  "target_title": "Orphée"
}
```

TABLEAU 1. EXTRAIT D'UNE CORRESPONDANCE DÉTECTÉE PAR TRACER

Id A	Id B	Similarité		Segment A	Segment B
		Nombre	%		
2600018	3200216	3	60	Un instant encore, et elles échappaient à la puissance du dieu des morts : elles touchaient au seuil du séjour des vivants.	Ainsi donc il touchait au seuil qui sépare l'empire des morts du séjour des vivants, lorsqu'il s'arrêta involontairement pour jeter un regard d'amour sur l'épouse qui lui était rendue, hélas ! rendue et ravie à-la-fois.

Hormis ce qui sert à identifier chaque extrait, aucun détail n'est fourni sur les entités qui motivent la mise en correspondance. Pourtant, notamment afin de pouvoir écarter les correspondances fausses ou peu pertinentes, nous devons savoir ce qui est apparié, les motivations des liens (identité de mots, de lemmes ou de racines, ou synonymes communs), le nombre d'entités communes et la distance qui les sépare. Pour pallier ces manques, nous post-traitions chacune des sorties, ce qui nous permet de les compiler et d'enrichir les résultats, de filtrer les couples peu ou pas pertinents et de générer une interface d'exploration offrant une vue globale et claire des couples de correspondances détectées.

Post-traitement des résultats des détections de correspondances

La chaîne de post-traitement, qui permet d'intégrer les résultats et les annotations enrichies à des fichiers XML dédiés, comporte trois étapes¹⁵ :

1. Unification des formats des résultats (compilation et XMLisation)
2. Analyse détaillée de chaque couple de correspondances afin de détecter les noyaux des syntagmes et les unités lexicales liés, leur nombre et leur degré de proximité, de calculer le score de ressemblance, puis d'écarter les couples trop faiblement liés
3. Adjunction d'appariements plus fins, notamment entre diverses dénominations d'entités nommées (étape non exposée ci-après)

Unification des formats (étape 1)

Nous retenons la segmentation du corpus en phrases associées à un identifiant unique, initialement réalisée pour Tracer. Quelques manipulations s'imposent pour découper ou réunir les segments détectés par TextPAIR selon le même principe. Ainsi, la correspondance potentielle présentée en figure 1 est révisée et étendue pour arriver au découpage du tableau 2 (où elle est en italique)¹⁶.

TABLEAU 2. IDENTIFICATION DES PHRASES CORRESPONDANTES DE BALLANCHE1809 ET BALLANCHE1829

BALLANCHE1809	BALLANCHE1829
<p>[18] Un instant encore, et elles échappaient à la puissance du dieu des morts : elles <i>touchaient au seuil du séjour des vivants</i>.</p> <p>[19] <i>Mais, ô faiblesse d'un cœur qui aime !</i></p> <p>[20] <i>Orphée s'arrête pour écouter le soupir</i> qui errait sur les lèvres d'Eurydice, pour prêter l'oreille au léger frôlement de ses vêtements aériens.</p>	<p>[216] Ainsi donc il <i>touchait au seuil qui sépare l'empire des morts du séjour des vivants, lorsqu'il s'arrêta involontairement</i> pour jeter un regard d'amour sur l'épouse qui lui était rendue, hélas ! rendue et ravie à-la-fois.</p>

Cependant, le traitement du texte seul, opéré par TextPAIR et Tracer, n'est pas toujours suffisant. Des indices précieux mais implicites dans les textes sont explicitement fournis par leur balisage en XML. Une annotation à plusieurs niveaux est ainsi associée à chaque phrase, syntagme et mot (figure 2, pour une partie de la phrase 21 de BALLANCHE1809¹⁷).

L'élément <s> porte deux attributs : @id qui indique son identifiant et @corresp qui fournit ceux des phrases détectées comme correspondantes. Les syntagmes nominaux, verbaux et adjectivaux sont balisés comme <phr> avec un @select qui donne le lemme du noyau lexical. Le syntagme nominal « le visage de son épouse » en contient un autre, « son épouse »¹⁸, aussi annoté comme étant une désignation de l'entité nommée « Eurydice » à l'aide de son élément parent <persName>, qui donne l'identité du référent avec un @corresp et l'identifiant de la périphrase avec un @id¹⁹. Les éléments <w> portent les codes grammaticaux (@pos) et flexionnels (@msd), les lemmes (@lemma) et, pour les unités lexicales pleines, une sélection de synonymes (@sameAs)²⁰.

FIGURE 2. BALISAGE D'UN EXTRAIT DE LA PHRASE 21 DE BALLANCHE1809

```
<s id="2600021" corresp="3000018 1300012 3201014 5600011 5900015"> [...]
  <phr type="GN" select="visage">
    <w id="2600021_47" n="47" lemma="le" pos="DET:ART">le</w>
    <w id="2600021_48" n="48" lemma="visage" pos="NOM" msd="ms"
      sameAs="figure physionomie tête air">visage</w>
    <w id="2600021_49" n="49" lemma="de" pos="PRP">de</w>
    <persName id="2600021_50-51" corresp="eurydice" type="périphrase">
      <phr type="GN" select="épouse">
        <w id="2600021_50" n="50" lemma="son" pos="DET:POS">son</w>
        <w id="2600021_51" n="51" lemma="épouse" pos="NOM" msd="fs"
          sameAs="conjointe bourgeoise femme régulière">épouse</w>
      </phr>
    </persName>
  </phr>
</s>
```

Appariement des noyaux de syntagmes (étape 2)

La recherche de relations lexicales au sein de chaque correspondance implique une comparaison qui exploite les valeurs des attributs en effectuant des appariements sur plusieurs niveaux.

Les résultats de Tracer pour la phrase 21 de BALLANCHE1809 montrent 12 couples impliquant 10 textes, dont 6 traductions de Virgile (4 en prose et 2 en vers) et 3 d'Ovide (en prose). Une partie de ces résultats se trouve sur la figure 3.

FIGURE 3. TROIS PHRASES EN CORRESPONDANCE AVEC LA PHRASE 21 DE BALLANCHE1809

	BALLANCHE1809, PHRASE 21	BALLANCHE1829, PHRASE 1014 [A]
☒ Noyaux lemmatisés de syntagmes ☐ [Noyaux identiques] ☐ [Noyaux synonymes] ☐ [Noyaux en relation flexionnelle ou dérivationnelle]	B[Vaincu]vaincre par cette puissance contre laquelle l'homme lutte C[en vain]vain, il B[C]se retourne]retourner et, B[soubliant]soublier B[sa fatale promesse]promesse, il permet à AB[son regard]regard d'C[interroger]interroger à la faveur de AB[la clarté naissante]clarté, A[les plis du A[voile]voile]pli qui lui cachait A[le visage de A[son épouse]épouse]visage, de cette touchante victime qu'il avait tant pleurée.	D'une main tremblante, celle qui veut être initiée reçoit des mains du poète [le voile tissu par [l'épouse d'un Titan]épouse]voile ; ce voile, emblème d'une science cosmogonique, descend à [longs plis]pli sur [le visage de la vierge]visage ; et [ses yeux]œil ont peine encore à soutenir la [clarté du jour]clarté. CABARET-DUPATY1897, PHRASE 11 [B] Presque aux portes du [jour]jour, il s'arrête, hélas ! il [oublie]oublier [sa promesse]promesse, et, [vaincu par l'amour]vaincre, il [tourne la tête]tourner pour [regarder]regarder sa chère Eurydice. CHARPENTIER1831, PHRASE 18 [C] [En vain]vain Orphée veut saisir son ombre fugitive ; en vain il [la rappelle]rappeler Eurydice ne revit plus Orphée et le sévère nocher ne lui permet plus de [repasser]repasser l'onde fatale.

Tout d'abord, seuls les lemmes noyaux de syntagmes sont pris en compte. Pour que cette correspondance nucléaire soit considérée comme effective, une équivalence exacte ou au niveau des synonymes est

exigée pour 30 % d'entre eux²¹.

26 La phrase 21 de BALLANCHE1809 compte 20 syntagmes (parfois enchâssés), donc le seuil est de 6. Pour BALLANCHE1829, nous trouvons 5 noyaux identiques et 1 équivalence synonymique. Les deux phrases étant de longueurs comparables, le ratio obtenu est minimal, ce qui attire l'attention sur la complexité de la reformulation entre les deux extraits. Par ailleurs, le noyau « épouse » est problématique : dans BALLANCHE1809, il renvoie à Eurydice, alors que dans BALLANCHE1829 la périphrase « l'épouse d'un Titan » désigne sa mère²².

27 La phrase de CABARET-DUPATY1897 (Virgile 1897 [trad. Cabaret-Dupaty]) compte 3 noyaux identiques, 1 équivalence synonymique et 2 radicaux correspondants, soit 6 syntagmes sur 11 appariés. Comme cette phrase est courte, le ratio atteint 55 % environ. Mais trois groupes importants de CABARET-DUPATY1897 restent isolés :

- « il s'arrête », que nous retrouvons dans la phrase 20 de BALLANCHE1809
- « l'amour », fortement métaphorisé par « cette puissance contre laquelle l'homme lutte en vain »
- et « sa chère Eurydice », périphrasé par « son épouse » et « cette touchante victime qu'il avait tant pleurée »

28 Le seuil de 30 % de correspondances nucléaires est atteint pour ces deux couples de phrases dont le traitement se poursuit.

29 Ce n'est pas le cas pour la correspondance avec CHARPENTIER1831 (Virgile 1831 [trad. Charpentier]) qui compte 1 locution adverbiale commune et 2 synonymes de verbes très polysémiques, soit 3 groupes appariés sur 11 (le ratio est de 27 %).

Recherche des relations lexicales (étape 2)

30 Pour tous les couples ayant atteint le seuil de correspondances nucléaires minimal, nous procédons à une comparaison lexicale, en écartant les mots faibles et en calculant un score de proximité lexicale. Pour le moment, trois types de relations sont pris en compte :

1. Mot à mot (forme~forme, lemme~lemme, radical~radical) : + 5 points
2. Entre synonymes (synonyme~synonyme) : + 0,5 point
3. Mixtes (mot~synonyme) : + 1,5 point

31 Quand une correspondance de type 1 est trouvée, les unités lexicales du couple ne sont plus prises en compte, mais pour les relations plus faibles la recherche se poursuit afin de trouver le couple lexical obtenant le meilleur score, et les points se cumulent si plusieurs correspondances entre synonymes ou mixtes sont trouvées. Comme 4 synonymes maximum sont associés aux mots pleins, le score lexical maximal (hors relation de type 1) peut atteindre 3 points : 1 correspondance mixte et 3 correspondances entre synonymes. L'équivalence entre « regard » et « œil », par exemple, obtient 2 points : le lemme « regard » figure parmi les synonymes d'« œil » et le synonyme « prunelle » est commun aux deux mots.

- 32 Les points de chaque couple lexical sont ensuite additionnés afin d'obtenir le score total des correspondances et de décider si le couple de phrases peut être retenu dans les résultats finaux.
- 33 Pour les couples de la figure 3, le traitement lexical ne révèle pas d'autres appariements que ceux déjà détectés. Ainsi, la correspondance avec BALLANCHE1829 obtient 27 points (6 relations forme~forme et 1 relation mixte) et celle avec CABARET-DUPATY1897 obtient 20,5 points (1 relation forme~forme, 2 relations lemme~lemme, 2 radical~radical, 1 synonyme~synonyme).
- 34 Nous calculons ensuite un ratio de ressemblance afin de pouvoir comparer des phrases de longueur et de complexité différentes. Le pourcentage du score obtenu est évalué par rapport au score maximal que le couple aurait si tous les mots forts de la phrase la plus courte entraient dans une relation mot à mot. Pour BALLANCHE1829, ce score lexical s'élève à 115 points (soit un ratio de 23,47 %), alors que pour la phrase plus courte de CABARET-DUPATY1897 il n'est que de 65 points (soit un ratio de 31,54 %).

Traitement des extraits étendus (étape 2)

- 35 Pour les résultats de TextPAIR en chevauchement sur plusieurs phrases, les relations entre ce qui correspond à chaque phrase peuvent être plus faibles et sont souvent insuffisantes pour atteindre les seuils imposés, alors que le couple détecté est pertinent dans sa totalité. Nous différencions donc le post-traitement des résultats des deux logiciels en conservant les mêmes seuils, tant au niveau des syntagmes que des unités lexicales, mais en effectuant la comparaison sur la totalité de l'extrait détecté par TextPAIR et non sur les phrases qui le composent.
- 36 TextPAIR a détecté une correspondance (déjà en partie présentée dans le tableau 2) qui regroupe 5 phrases de BALLANCHE1809 (17-21) et 3 de BALLANCHE1829 (215-217). Tracer n'a retenu que le couple des phrases 17 et 215. Or, la correspondance (figure 4) est valide pour la totalité de l'extrait : le ratio nucléaire s'élève à 42,6 % (23 sur 54) et le ratio lexical à 37 % (106 points sur 285). Les entités communes ne se répartissent pas équitablement : la phrase 21, par exemple, a 1 entité commune avec la phrase 215, 3 avec la 216 et 2 avec la 217.

FIGURE 4. CORRESPONDANCES LEXICALES ENTRE BALLANCHE1809 (17-21) ET BALLANCHE1829 (215-217)

CHOOSE LA CORRESPONDANCE
ballanche1829, ph. 215-217

☒ Relations lexicales

- forme-forme
- lemme-lemme
- radical-radical
- synonyme-synonyme
- ou mixte

BALLANCHE1809, PHRASES 17-21

[17] Déjà¹ les ombres² devenaient³ moins opaques ; déjà un faible⁴ crépuscule⁵ détache⁶ des rayons⁷ du soleil⁸, arrivait jusqu'aux yeux⁹ des deux nobles créatures qui s'avançaient vers la lumière¹⁰ du jour¹¹.

[18] Un instant encore, et elles échappaient à la puissance¹² du dieu des morts¹³ : elles touchaient¹⁴ au seuil¹⁵ du séjour¹⁶ des vivants¹⁷.

[19] Mais, ô faiblesse d'un cœur¹⁸ qui aime !

[20] Orphée s'arrête¹⁹ pour écouter le soupir qui errait sur les lèvres d'Eurydice²⁰, pour prêter l'oreille au léger frôlement de ses vêtements aériens.

[21] Vaincu par cette puissance contre laquelle l'homme lutte en vain, il se retourne²¹ ; et, oubliant²² sa fatale promesse, il permet à son regard²³ d'interroger, à la faveur de la clarté naissante, les plis du voile²⁴ qui lui cachait le visage²⁵ de son épouse²⁶, de cette touchante victime qu'il avait tant pleurée.

BALLANCHE1829, PHRASES 215-217

[215] Déjà¹ un crépuscule⁵ douteux⁴ commençait à remplacer l'obscurité² immense ; un rayon de pâle lumière¹⁰, qui vint flotter autour de ses yeux⁹, lui fit oublier²² le passé et l'avenir : ce n'était pas une lueur⁸ trompeuse, c'était bien un rayon⁷ détaché⁶ de la clarté du jour¹¹ ; la marque de son salut devint³ le signe de sa perte.

[216] Ainsi donc il touchait¹⁴ au seuil¹⁵ qui sépare l'empire¹² des morts¹³ du séjour¹⁶ des vivants¹⁷, lorsqu'il s'arrêta¹⁹ involontairement pour jeter un regard²³ d'amour¹⁸ sur l'épouse²⁶ qui lui était rendue²¹, hélas ! rendue et ravie à-la-fois.

[217] Si, du moins, il eût eu le temps de soulever le voile²⁴ qui retombait sur le visage²⁵ d'Eurydice²⁰ !

37 Ainsi, l'exploration des correspondances, tant nucléaire que lexicale, permet, d'un côté, l'exclusion des relations fausses ou peu pertinentes et, de l'autre, l'intégration et la validation des réécritures effectuées sur de longs fragments, même en cas de reformulations complexes et non linéaires.

38 Ces manipulations fournissent aussi de précieuses informations pour l'analyse littéraire et linguistique des correspondances validées automatiquement. Les exemples des figures 3 et 4 révèlent principalement deux types de problèmes :

- Les périphrases des entités nommées ne sont appariées ni entre elles ni avec les noms propres de leur référent, même quand il s'agit d'Eurydice, qui est pourtant souvent évoquée. L'exploitation des attributs de ces éléments très fréquents permet d'éviter les appariements fautifs ou, au contraire, d'en détecter d'autres.
- Certains appariements de synonymes très polysémiques (principalement des verbes fréquents) sont pertinents pour certains de leurs emplois, mais pas pour toutes leurs constructions ni pour les valeurs sémantiques attestées dans notre corpus. Le verbe « retourner » employé pronominalement, par exemple, fait partie des mots-clés saillants de l'épisode de la seconde mort d'Eurydice, mais son sens dans ce contexte est bien précis et ne correspond pas aux synonymes les plus courants du verbe en langue générale (toutes constructions confondues), comme « rendre », « repasser » ou « rappeler ».

39 Ces difficultés peuvent être surmontées grâce à l'élaboration de lexiques adaptés à la spécificité du corpus couplés à des dictionnaires de langue générale (notamment de synonymes) et à des ressources polyvalentes (utilisées pour le repérage de chaînes de références). Ces lexiques sont progressivement enrichis grâce aux détections de correspondances validées.

Exploitation des résultats enrichis

40 Les appariements (poly-)lexicaux validés sont intégrés dans le balisage des phrases. Pour l'extrait de la phrase 21 de BALLANCHE1809 (figure 2), deux relations exactes sont établies avec le mot « épouse ». Elles sont signalées (figure 5) à l'aide de l'élément <xr>, enfant de <w>, qui renseigne l'identifiant du mot correspondant (@corresp), le type de la relation détectée (@type) et le score obtenu par le couple (@cert). L'annotation des périphrases et des noms propres d'entités ajoute, elle, plusieurs correspondances à « son épouse », qui trouve des appariements pertinents dans 5 textes sur 6. Le lien fautif avec la périphrase pour la mère d'Eurydice « l'épouse d'un Titan » chez BALLANCHE1829 est remplacé par un lien avec « la vierge²³ » et 1 correspondance est ajoutée avec « sa chère Eurydice » de CABARET-DUPATY1897.

```

<persName id="2600021_50-51" corresp="eurydice" type="périphrase">
  <xr corresp="3200216_30-35" type="périphrase-périphrase" cert="5"/>
  <xr corresp="1300012_9-10" type="périphrase-périphrase" cert="5"/>
  <xr corresp="3201014_6-10" type="périphrase-périphrase" cert="5"/>
  <xr corresp="5600011_31-33" type="périphrase-entité-nommée" cert="5"/>
  <xr corresp="5900015_63-64" type="périphrase-entité-nommée" cert="5"/>
  <phr type="GN" select="épouse">
    <w id="2600021_50" n="50" lemma="son" pos="DET:POS">son</w>
    <w id="2600021_51" n="51" lemma="épouse" pos="NOM" msd="fs"
      sameAs="conjointe bourgeoise femme régulière">épouse
      <xr corresp="3200216_31" type="forme-forme" cert="5"/>
      <xr corresp="1300012_10" type="forme-forme" cert="5"/>
    </w>
  </phr>
</persName>

```

41 En fin de traitement, le fichier XML de chaque texte est significativement enrichi et une base XML compile les correspondances valides détectées par TextPAIR et Tracer. Ce double enregistrement des résultats permet un large éventail de manipulations et d'analyses, tantôt spécifiques, concentrées sur un texte précis, tantôt transversales, incluant tout ou partie des textes.

Exploration des correspondances à l'aide de graphes

42 Pour la phrase 21 de BALLANCHE1809, par exemple, un nombre assez important de correspondances s'observe avec les différents textes du corpus. Pour mieux exploiter ce type de relations complexes, une visualisation des données avec des graphes de correspondances, au sein desquels la taille d'un nœud renvoie au nombre de phrases qui composent l'extrait et l'épaisseur d'un lien reflète le ratio de proximité lexicale entre chaque couple (figure 6), se révèle particulièrement adaptée²⁴.

FIGURE 6. GRAPHE DE CORRESPONDANCES DE LA PHRASE 21 DE BALLANCHE1809

43 La phrase 21 est intégrée dans un ensemble plus large, qui regroupe 6 phrases (de 16 à 21), grâce, d'une part, à la correspondance avec BALLANCHE1829 (figure 4) et, d'autre part, à une correspondance des phrases 16 et 17 avec le poème *Orphée* (Launay 1922 : LAUNAY1922 dans le corpus). Sept relations au total ont été établies :

- 3 avec des traductions de Virgile (1846 [trad. Desportes : DESPORTES1846_PROSE], 1863 [trad. Guillois : GUILLOIS1863], 1897 [trad. Cabaret-Dupaty : CABARET-DUPATY1897])
- 1 avec une traduction d'Ovide (1880 [trad. Parnajon : PARNAJON1880_VERS])

- 1 avec la transmodalisation de Launay (1922)
- 2 autres, possibles, avec des extraits très éloignés de BALLANCHE1829 : les phrases 215-217, d'un côté, et la phrase 1014, de l'autre

44

En effet, l'extrait BALLANCHE1829 (215-217) fait partie d'un chant du fils du roi Évandre, qui raconte la version originelle du mythe, rectifiée ensuite par le poète Thamyris²⁵. Concernant Eurydice, cette rectification se retrouve au sein du second extrait (1014), où elle meurt pour toujours au moment de l'accomplissement du rite de son mariage avec Orphée. Contrairement au récit de Virgile, cette disparition n'a pas lieu aux Enfers et n'est pas précédée d'une plainte de l'épouse²⁶. À la place, Eurydice s'excuse²⁷ d'avoir distrahit le poète dans sa mission de fondateur des cultes et l'encourage à la poursuivre, malgré la douleur de la perte qui devient alors initiatique²⁸. Si nous nous penchons sur la totalité des correspondances détectées pour BALLANCHE1829, nous nous rendons compte qu'il s'agit de l'unique relation permettant de lier les deux descriptions de la mort d'Eurydice entre elles et avec les sources anciennes (figure 7).

FIGURE 7. GRAPHE DE CORRESPONDANCES DES PHRASES 215-217 ET 1014 DE BALLANCHE1829

45

Un réseau complexe peut en revanche être observé à proximité de l'extrait 215-217 (figure 8) : pour les phrases 219-220, 10 correspondances ont été trouvées avec les traductions de Virgile et 1 avec le poème LAUNAY1922.

46

L'essai BALLANCHE1809 s'arrête à la seconde mort d'Eurydice, il n'y a donc pas d'autres relations avec cette partie de l'épopée (BALLANCHE1829). Mais il y a beaucoup de correspondances entre cette dernière et les traductions dans la suite de la description du mythe. Toutefois, l'auteur ne procède pas de la même manière pour la deuxième description de la mort d'Eurydice, beaucoup plus allusive, ce qui empêche la détection de correspondances autres que celle avec BALLANCHE1809.

47

Enfin, des éléments de la conclusion de l'essai, dans laquelle l'interprétation du mythe est développée, peuvent également être mis en correspondance avec deux extraits de BALLANCHE1829²⁹. Pour BALLANCHE1809, la conclusion renvoie clairement à l'épisode de la mort d'Eurydice, mais, dans l'épopée, les correspondances font partie d'une description de cérémonie funèbre en Égypte ancienne et de la métaphore de la vie de l'homme³⁰. Cette observation permet de relativiser ou d'appuyer quelques analyses traditionnelles du *Sixième Fragment*. Paulson (1986) procède à l'interprétation biographique de la rupture amoureuse de Ballanche, en comparant la remontée des Enfers au « temps bref des illusions de l'amour » (25). Pour Juden (1970), la douleur de la mort d'Eurydice « résume le sort de l'humanité, séparée de l'unité » (142). La prise en compte de la partie interprétative de l'essai dans le contexte des correspondances trouvées avec l'épopée confirme que la figure d'Eurydice comme le deuil de sa mort dépassent la symbolique amoureuse. La conclusion fait écho, dans l'épopée, à la description de la mort et de l'errance de l'âme à la recherche du salut, mais aussi à l'expression de la condition des humains, qui transmettent la souffrance avec le savoir. Le *Sixième Fragment* est majoritairement analysé dans le contexte des autres essais de Ballanche et de sa biographie, ou considéré comme une préfiguration de l'épopée, manquant peut-être encore de la profondeur métaphysique que l'auteur donnera au mythe 20 ans plus tard. Or, comme nous espérons l'avoir démontré, la proximité effective entre les deux textes, matérialisée par des allusions et d'importantes reformulations, prouve que, dans l'essai, la réflexion de Ballanche autour du mythe d'Orphée est déjà bien entamée.

Exploration du réseau global des réécritures

48

La vue globale des résultats, enregistrés dans la base XML, se révèle, quant à elle, particulièrement utile pour l'observation des relations que les œuvres entretiennent. Pour BALLANCHE1829, par exemple, des correspondances ont été détectées avec 65 textes du corpus, dont 34 réécritures (figure 9³¹). Toutefois, ce texte est très long et le récit de l'amour avec Eurydice y est ponctuel et rédigé de manière très figurée. Ces facteurs importent dans l'analyse de la place de BALLANCHE1829 dans le réseau des relations. Ainsi, au lieu de décompter le nombre des correspondances détectées pour chaque œuvre, il est plus pertinent de comparer les différents scores obtenus et leurs moyennes. Par exemple, 8 correspondances seulement ont été trouvées avec BALLANCHE1809, mais la moyenne du ratio lexical s'élève à 27 %, le plaçant ainsi dans les 10 textes les plus proches alors que celle des 42 correspondances avec LAUNAY1922 n'est que de 17,45 %, moins que la moyenne générale de la totalité des résultats (19 %).

49

Parmi les réécritures les plus proches, nous trouvons deux textes en prose, *Histoire des amants célèbres de l'antiquité* (Esquiros et Esquiros 1849) et *Les Grands Initiés. Esquisse de l'histoire secrète des religions* (Schuré 1889). Pour eux, la correspondance avec BALLANCHE1829 compte le plus de correspondances (16 et 12) et se place systématiquement dans les 10 textes les plus proches, peu importe le critère de comparaison. Ces œuvres insistent sur le rôle d'Orphée comme fondateur des cultes. Chez Esquiros et Esquiros, la catabase est remplacée par des épreuves initiatiques, auxquelles Orphée est soumis par des prêtres de l'Égypte ancienne. Eurydice, mordue par le serpent, est seulement endormie. Schuré, lui, met en scène un Orphée déjà initié, transmettant son savoir à des disciples. Dans cette réécriture, Eurydice meurt empoisonnée par Aglaonice de Thessalie, une bacchante, prêtresse d'Hécate et « éprise pour [Eurydice] d'une envie perverse, d'un amour effréné, maléfique » (258). Dans la pièce de théâtre *Orphée* (Cocteau 1927), Aglaonice endosse un rôle comparable, mais fortement caricaturé³². Des relations très fortes (dignes d'une reprise : 6 correspondances pour un ratio de 78 %) ont été détectées entre Schuré (1889) et *Orphée d'Afrique, théâtre-rituel* (Ma Njock 2000), une adaptation du roman *Orphée-Dafric* (Liking 2000), qui transfère le mythe sur le continent africain, dans la continuité d'un mythe plus récent, celui de l'« Orphée noir »³³. Porra (2012) constate chez Liking et Ma Njock l'influence conjointe de textes européens et de la réalité africaine, et note l'importance de l'aspect initiatique, focalisé autour d'une lecture mythique égyptienne (§ 19-27), mais ne relève pas de relation forte avec Schuré (1889). Un examen plus approfondi s'impose pour confirmer l'absence de lien dégagé par les recherches traditionnelles, mais elle est plausible compte tenu des convictions de Schuré³⁴.

FIGURE 9. RÉSEAU DES RELATIONS DE BALLANCHE1829

50 S'il est satisfaisant d'observer le réseau de relations, qui se construit à partir des correspondances détectées, le plus souvent en confirmant ce qui a déjà pu être observé, parfois en révélant de nouveaux liens, il convient d'admettre quelques manques importants à ce stade de l'étude. Par exemple, Brunel (2016) relève plusieurs épisodes inspirés par Schuré (1889) dans la pièce de théâtre symbolique *Orphée-Roi* (Segalen 1921). Or, au sein de notre corpus, une seule correspondance a été trouvée – et encore, elle l'a été grâce aux didascalies qui introduisent le premier acte. Elle a l'intérêt de permettre d'identifier deux personnages anonymes, mais les proximités relevées par Brunel concernent des épisodes beaucoup plus significatifs, comme celui de la lyre ou de la mort d'Eurydice.

51 Cette absence de détection de correspondances vient du fait qu'au sein des œuvres dramatiques, l'enchaînement de répliques courtes accompagnées de didascalies impose soit de cumuler des segments très courts au sein desquels peu de correspondances sont trouvées, soit de ne pas segmenter et de générer des détections fautives. Les didascalies sont, pour la plupart, peu pertinentes, notamment quand elles désignent les personnages ou des mouvements scéniques. Mais, comme dans *Orphée-Roi*, elles véhiculent parfois l'indice de réécriture.

52 Nous espérons que, à l'avenir, notre traitement saura prendre en compte ces éléments structurels et intégrera pleinement les œuvres dramatiques au sein du réseau des réécritures. Cependant, nous n'ambitionnons pas de supplanter les analyses empiriques des chercheurs traditionnels qui, comme Brunel, comparent deux textes en prenant en compte la totalité des œuvres d'un auteur, y ajoutent leur sensibilité littéraire et trouvent parfois un sens subtil même à partir d'un mot qui semble, à première vue, insignifiant. La plus-value d'une recherche appuyée sur les outils numériques réside, à notre sens, dans leur capacité de traiter simultanément de larges ensembles de textes et d'attirer l'attention du chercheur sur des phénomènes peu visibles.

Conclusion

53 Ainsi, si la possibilité de détecter automatiquement des relations intertextuelles est passionnante, pour la recherche littéraire et linguistique et pour l'édition (numérique), les détections fautives et les problèmes de reconnaissance inhérents confirment la nécessité de combiner des traitements (semi-automatiques) et des méthodes non informatiquement outillées. Une connaissance approfondie du corpus est par ailleurs nécessaire, tant pour analyser les résultats obtenus que pour les améliorer et les enrichir, afin que les traitements réitérés des mêmes textes bénéficient des premières passes et deviennent plus fructueux et plus aptes à dévoiler de nouvelles relations, de plus en plus fines et peu visibles.

54 En partant d'une œuvre particulière, l'essai de Ballanche, et des correspondances trouvées entre elle et d'autres textes du corpus, nous avons analysé les résultats des détections de correspondances opérées par les logiciels TextPAIR et Tracer, montré la finesse et l'importance des post-traitements et permis de confronter les relations dégagées et interprétées aux analyses de facture classique du topos orphique.

55 Comme les recherches traditionnelles, celles soutenues par des outils informatiques exigent un travail minutieux avant d'obtenir des résultats probants. Les deux méthodes ne sont pas concurrentes, elles seront certainement d'autant plus efficaces qu'elles seront combinées.

Bibliographie

Allen, Timothy, Charles Cooney, Stéphane Douard, Russell Horton, Robert Morrissey, Mark Olsen, Glenn Roe et Robert Voyer. 2010. « Plundering Philosophers : Identifying Sources of the *Encyclopédie* ». *Journal of the Association for History and Computing* 13 (1). <http://hdl.handle.net/2027/spo.3310410.0013.107>.

Ballanche, Pierre-Simon. 1801. *Du sentiment considéré dans ses rapports avec la littérature et les arts*. Lyon, Paris : Ballanche et Barret, Calixte Volland.

Ballanche, Pierre-Simon. 1809. *Sixième Fragment*. Dans *Œuvres de M. Ballanche*, 488-492. Paris : J. Barbezat.

Ballanche, Pierre-Simon. 1819. *Antigone*. Paris : Jules Didot.

Ballanche, Pierre-Simon. 1827. *Essais de palingénésie sociale*. *Orphée*. Paris : Jules Didot.

Barzilay, Regina et Kathleen R. McKeown. 2001. « Extracting Paraphrases from a Parallel Corpus ». Dans *Proceedings of the 39th Annual Meeting of the ACL*, 50-57. Toulouse : Association for Computational Linguistics. <https://doi.org/10.3115/1073012.1073020>.

Béague, Annick, Jacques Boulogne, Alain Deremetz et Françoise Toulze. 2019. *Les Visages d'Orphée*. Villeneuve d'Ascq : Presses universitaires du Septentrion. <https://doi.org/10.4000/books.septentrion.50854>.

Bergroth, Laito, Harri Hakonen et Timo Raita. 2000. « A Survey of Longest Common Subsequence Algorithms ». Dans *Proceedings of the Seventh International Symposium on String Processing Information Retrieval*, 39-48. IEEE. <https://doi.org/10.1109/SPIRE.2000.878178>.

Boukhaled, Mohamed-Amine, Zied Sellami et Jean-Gabriel Ganascia. 2015. « Phoebus : un logiciel d'extraction de réutilisations dans des textes littéraires ». Communication présentée à la 22^e *Conférence sur le traitement automatique des langues naturelles*, Caen, 22-25 juin. <https://hal.sorbonne-universite.fr/hal-01198411>.

Bouloumié, Arlette. 2004. « La résurgence du mythe d'Eurydice et ses métamorphoses dans l'œuvre d'Anouilh, de Pascal Quignard, d'Henri Bosco, de Marguerite Yourcenar, de Michèle Sarde, et Jean Loup Trassard ». *Loxias* 2. <http://revel.unice.fr/loxias/index.html?id=1244>.

Brunel, Pierre. 1997. « Orphée moderne ». Dans *Apollinaire entre deux mondes. Le Contre-point mythique dans Alcools. Mythocritique II*, 63-82. Paris : Presses universitaires de France.

Brunel, Pierre. 2016. « Orphée-roi de Victor Segalen, ou le miracle de la lyre ». Dans *Mythocritique. Théorie et parcours*, 165-174. Grenoble : UGA Éditions. <https://doi.org/10.4000/books.ugaeditions.6498>.

Büchler, Marco. 2013. « Informationstechnische Aspekte des Historical Text Re-use ». Leipzig : université de Leipzig.

Cocteau, Jean. 1927. *Orphée. Tragédie en un acte et un intervalle*. Paris : Stock.

Coffee, Neil, Jean-Pierre Koenig, Shakthi Poornima, Roelant Ossewaarde, Christopher Forstall et Sarah Jacobson. 2012a. « Intertextuality in the Digital Age ». *Transactions of the American Philological Association* 142 (2) : 383-422. <https://doi.org/10.1353/apa.2012.0010>.

Coffee, Neil, Jean-Pierre Koenig, Shakthi Poornima, Christopher Forstall, Roelant Ossewaarde et Sarah Jacobson. 2012b. « The Tesseract Project : Intertextual Analysis of Latin Poetry ». *Literary and Linguistic Computing* 28 (2) : 221-228. <https://doi.org/10.1093/lc/fqs033>.

Del Lungo, Andrea et Karolina Suchecka. [À paraître]. « Projet eBalzac : construire une bibliothèque hypertextuelle des sources intertextuelles ». Dans *La Fabrique numérique des corpus en sciences humaines et sociales*, édité par Clarisse Bardiot, Esther Dehoux et Émilien Ruiz. Villeneuve d'Ascq : Presses universitaires du Septentrion.

Deshayes, Prosper-Didier et Jean Rouhier-Deschamps. 1792. *Le Petit Orphée : opéra comique*. Paris : Tourbon.

Esquiros, Alphonse et Adèle Esquiros. 1849. *Histoire des amants célèbres de l'antiquité*. Paris : Bureau de la Société fraternelle de l'intelligence et du travail réunis.

Forstall, Christopher, Neil Coffee, Thomas Buck, Katherine Roache et Sarah Jacobson. 2015. « Modeling the Scholars : Detecting Intertextuality through Enhanced Word-Level N-Gram Matching ». *Digital Scholarship in the Humanities* 30 (4) : 503-515. <https://doi.org/10.1093/lc/fqu014>.

Franzini, Greta. 2016. « English Translations of *Pan Tadeusz* : a Comparison with TRACER ». *eTRAP* (blog). <https://www.etrapp.eu/english-translations-of-pan-tadeusz-a-comparison-with-tracer/>.

Ganascia, Jean-Gabriel. 2021. « Détection automatique de phénomènes intertextuels », *Genesis* 51 : 63-77. <https://doi.org/10.4000/genesis.5671>.

Ganascia, Jean-Gabriel, Pierre Glaudes et Andrea Del Lungo. 2014. « Automatic Detection of Reuses and Citations in Literary Texts ». *Literary and Linguistic Computing* 29 (3) : 412-421. <https://doi.org/10.1093/lc/fqu020>.

Gladstone, Clovis. 2018. « Evaluating the Practices and Legacy of the Enlightenment on 19th-Century Print Culture ». *ARTFL* (blog). <https://artfl.blogspot.com/2018/11/evaluating-practices-and-legacy-of.html>.

Gluck, Christoph Willibald, Ranieri de Calzabigi et Pierre-Louis Moline. 1774. *Orphée et Eurydice*. Paris : Académie Royale.

Gusfield, Dan. 1997. *Algorithms on Strings, Trees, and Sequences : Computer Science and Computational Biology*. Cambridge : Cambridge University Press. <https://doi.org/10.1017/CBO9780511574931>.

Gyssels, Kathleen. 2005. « Sartre postcolonial ? Relire “Orphée noir” plus d’un demi-siècle après ». *Cahiers d’études africaines* 45 (179-180) : 631-650. <https://doi.org/10.4000/etudes-africaines.14952>.

Horton, Russell, Mark Olsen et Glenn Roe. 2010. « Something Borrowed : Sequence Alignment and the Identification of Similar Passages in Large Text Collections ». *Digital Studies / Le Champ numérique* 2 (1). <https://doi.org/10.16995/DSCN.258>.

Jänicke, Stefan, Annette Gessner, Marco Büchler et Gerik Scheuermann. 2014. « Visualizations for Text Re-Use ». Dans *Proceedings of the 5th International Conference on Information Visualization Theory and Applications*, 59-70. Setubal : Science and Technology Publications. <https://doi.org/10.5220/0004692500590070>.

Jarrassé, Dominique. 2009. « Trois gouttes d’art nègre. Gobinisme et métissage en histoire de l’art ». *Les Actes de colloques du musée du quai Branly Jacques Chirac* 1 (juillet). <https://doi.org/10.4000/actesbrany.96>.

- Juden, Brian. 1970. « Particularités du mythe d'Orphée chez Ballanche ». *Cahiers de l'AIIEF* 22 (1) : 137-152. <https://doi.org/10.3406/caief.1970.956>.
- Jurafsky, Daniel et James H. Martin. 2009. « N-Gram Language Models ». Dans *Speech and Language Processing : An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 189-232. New York : Prentice-Hall, Inc.
- Kokkinakis, Dimitrios et Mats Malm. 2016. « Detecting Reuse of Biblical Quotes in Swedish 19th-Century Fiction using Sequence Alignment ». Dans *Proceedings of the Workshop on Corpus-based Research in the Humanities (CRH)*, édité par Francesco Mambrini, Marco Passarotti et Caroline Sporleder, 79-86. Varsovie, Pologne : Polish Academy of Sciences.
- Kushner, Eva. 1961. *Le Mythe d'Orphée dans la littérature française contemporaine*. Paris : A.G. Nizet.
- Launay, Louis de. 1922. *Orphée : poème*. Paris : Jouve et C^{ie} éditeurs.
- Liking, Werewere. 2000. *Orphée Dafric*. Paris : L'Harmattan.
- Ma Njock, Manuna. 2000. *Orphée d'Afrique, théâtre-rituel*. Paris : L'Harmattan.
- Monteverdi, Claudio et Alessandro Striggio. 1609. *L'Orfeo, favola in musica [Orphée, fable en musique]*. Venise : Ricciardo Amadino.
- Ovide. 1880. « Choix des Métamorphoses ». Dans *Les Auteurs latins expliqués d'après une méthode nouvelle par deux traductions françaises*, édité par une société de professeurs et de latinistes, traduit par Félix de Parnajon, 418-429, 440-447. Paris : Hachette.
- Paulson, William R. 1986. « Fragment et autobiographie dans l'œuvre de Ballanche : étude et textes inédits ». *Nineteenth-Century French Studies* 15 (1/2) : 14-32. <http://www.jstor.org/stable/23532414>.
- Porra, Véronique. 2012. « Sur quelques Orphée noirs. Reproduction, adaptation et hybridation du mythe d'Orphée en contexte post(-)colonial ». *Revue de littérature comparée* 344 (4) : 441-455. <https://doi.org/10.3917/rlc.344.0441>.
- Reboul, Marianne. 2017. « Comparaison semi-automatique des traductions en langue française de l'*Odyssée* d'Homère (1547-1955) ». Thèse de doctorat. Paris : université Paris IV.
- Riesz, János. 2005. « "Orphée noir" – "Schwarzer Orpheus" – "Black Orpheus" : quand les marges se mettent en marche vers un centre commun ». *Revue de littérature comparée* 314 (2) : 161-177. <https://doi.org/10.3917/rlc.314.0161>.
- Sartre, Jean-Paul. 1948. « Orphée noir ». Dans *Anthologie de la nouvelle poésie nègre et malgache de langue française*, édité par Léopold Sédar Senghor. Paris : Presses universitaires de France.
- Schuré, Édouard. 1889. « Livre V. Orphée (Les mystères de Dionysos) ». Dans *Les Grands Initiés. Esquisse de l'histoire secrète des religions*. 219-264. Paris : Perrin.
- Segalen, Victor. 1921. *Orphée-roi*. Paris : Georges Crès.
- Suchecka, Karolina et Nathalie Gasiglia. [À paraître]. « On Digital Comparative Edition and Textual Similarity Detection Tools : Towards a Hypertextual Cartography of a Rewritten Myth ». *Plotting Poetry 4. Tackling the Toolkit*.
- Suchecka, Karolina, Nathalie Gasiglia et Karl Zieger. 2019. « Édition comparative intermédiaire de séries traductives : exploiter les homologues pour créer des visualisations modulables ». *Traitement automatique des langues*, « TAL et humanités numériques », 60 (3) : 37-62. <https://www.atala.org/content/%C3%A9dition-comparative-interm%C3%A9diaire-de-s%C3%A9ries-traductives-exploiter-les-homologies-pour-cr%C3%A9er>.
- Virgile. 1831. « Géorgiques ». Dans *Œuvres complètes de Virgile, traduction nouvelle*, traduit par Jean-Pierre Charpentier, 1 : 259-265. Paris : Panckoucke.
- Virgile. 1846. « Quatrième livre ». Dans *Géorgiques*, traduit par Auguste Desportes, 49-57. Paris : Hachette.
- Virgile. 1863. « Quatrième livre ». Dans *Géorgiques*, traduit par « un naturaliste », 65-68. Paris : impr. de V^{ve} Guillois.
- Virgile. 1897. « Géorgiques. Livre IV ». Dans *Œuvres complètes de Virgile*, traduit par J.-R.-T. Cabaret-Dupaty, 101-104. Paris : Hachette.
- Wise, Michael J. 1995. « Neweyes : A System for Comparing Biological Sequences Using the Running Karp-Rabin Greedy String-Tiling Algorithm ». Dans *ISMB-95 Proceedings*, 393-401, AAAI. <https://www.aaai.org/Papers/ISMB/1995/ISMB95-047.pdf>.

Notes

1 Voir par exemple Coffee *et al.* 2012a ; Barzilay et McKeown 2001 ; Ganascia, Glaudes et Del Lungo 2014.

2 Voir <https://artfl-project.uchicago.edu/text-pair/> (ARTFL Project, université de Chicago) et Horton, Olsen et Roe 2010.

3 Voir <https://www.etrapp.eu/research/tracer/> (projet eTRAP, université de Göttingen) et Büchler 2013.

4 Voir par exemple Wise 1995 ; Gusfield 1997 ; Bergroth, Hakonen et Raita 2000.

5 Les textes libres de droits de notre corpus sont disponibles au format XML-TEI sur <https://github.com/karolinasucheka/orphidys/>. Nous partageons également nos scripts, que nous présentons rapidement dans cet article, et les résultats enrichis. Notez toutefois que nous ne sommes pas au stade final de notre recherche et que l'adaptabilité de nos développements n'a pas encore été testée. Nous serons heureuses de recevoir tout retour et commentaire quant à l'utilité des ressources mises à disposition.

6 Brunel étudie la présence d'Orphée principalement dans les œuvres poétiques et dramatiques du ^{xx}^e siècle, afin de démontrer le passage de sa figure archaïque aux représentations modernes. Ce faisant, il dégage plusieurs manières dont le mythe est repris (dans sa totalité ou par fragments) et établit cinq procédés de modernisation du mythe (18-19). Ces derniers induisent de réelles difficultés pour la détection outillée, alors qu'ils en posent moins pour des méthodes empiriques.

7 La présence du mythe chez Ballanche a été étudiée avec des méthodes empiriques traditionnelles (voir notamment Juden 1970 et Paulson 1986). Notre démonstration s'appuie sur les œuvres de cet auteur, mais n'aborde pas le sujet dans toute sa complexité.

8 Ou séquences des mots : voir Jurafsky et Martin 2009 et, pour la détection d'intertextualité, Forstall *et al.* 2015.

9 Ils sont conçus pour permettre de regrouper ceux du même texte : 0100001 pour le premier segment du premier texte, etc.

10 Voir par exemple Allen *et al.* 2010 et Gladstone 2018 pour TextPAIR, Franzini 2016 et Kokkinakis et Malm 2016 pour Tracer.

11 Voir par exemple Tesserae (Coffee *et al.* 2012b), Phœbus (Boukhaled, Sellami et Ganascia 2015) ou Odyseus (Reboul 2017).

12 Voir les *TEI Guidelines*, <https://tei-c.org/guidelines/>.

13 Les améliorations possibles sont, par exemple, la reprise des éléments facilitant la localisation de l'extrait détecté (<milestone> pour le numéro de page ou <s> pour l'identifiant de la phrase) ou la prise en compte de divers enrichissements, comme l'annotation et l'identification des noms propres et de leurs périphrases (<persName>, <ref>, etc.).

14 Pour nos traitements avec TextPAIR, 3 mots maximum peuvent séparer une entité commune de l'autre. Même avec un seuil de 5, les résultats pour des phrases longues ne sont pas aussi probants que ceux de Tracer, alors que le nombre de détections fautives augmente nettement.

15 Dans cet article, notre objectif principal est de démontrer le lien entre le post-traitement proposé et les analyses littéraires qu'il facilite. Pour l'évaluation, la validation et la discussion de notre méthode, voir Suchecka et Gasiglia à paraître.

16 Nous verrons plus loin, sous « Traitement des extraits étendus (étape 2) », que le couple de correspondances détecté par Tracer (tableau 1) peut en fait être intégré au sein d'un extrait plus large et que cela constitue un indice de la manière dont Ballanche se réécrit lui-même. Pour observer la correspondance dans sa totalité, il faudra l'étendre aux phrases 17-21 de BALLANCHE1809 et 215-217 de BALLANCHE1829.

17 « Vaincu par cette puissance contre laquelle l'homme lutte en vain, il se retourne ; et, oubliant sa fatale promesse, il permet à son regard d'interroger, à la faveur de la clarté naissante, les plis du voile qui lui cachait le visage de son épouse, de cette touchante victime qu'il avait tant pleurée. » (Ballanche 1809, 490.)

18 Il s'agit d'un syntagme prépositionnel introduit par « de », mais notre objectif étant de repérer des dénominations, seuls les syntagmes nominaux sont balisés, les syntagmes prépositionnels pouvant être restitués en leur associant la préposition antéposée.

19 La valeur de l'attribut @id est constituée de l'identifiant de la phrase, de « _ » et des rangs des premiers et derniers mots de la périphrase séparés par un tiret. Un @id est assigné à chaque mot selon le même schéma.

20 Ils sont 4, d'abord extraits d'un dictionnaire de synonymie cumulative de langue générale (*Dictionnaire électronique des synonymes*, Crisco, université de Caen, <https://crisco2.unicaen.fr/des/>), au sein duquel les corrélats sont classés par nombre décroissant de sens élémentaires communs avec la vedette (voir <http://crisco.unicaen.fr/dictionnaire-electronique-des-synonymes/presentation-du-des/>), puis remplacés, au fil des explorations, par des synonymes attestés dans le corpus (voir sous « Traitement des extraits étendus (étape 2) »).

21 Cette valeur dépend des paramètres utilisés pour le traitement. Pour un nombre minimal de mots communs fixé à 3 et une recherche par trigrammes, le seuil de 30 % semble être le plus adapté, car, à 25 %, beaucoup plus de fausses correspondances sont retenues, alors que, à 35 %, quelques détections pertinentes sont perdues.

22 Les sources antiques du mythe ne mentionnent pas les parents d'Eurydice, mais chez Ballanche, elle est la fille de Talaon, le dernier Titan, et de Vola, une nymphe.

23 Dans ces extraits, nous retrouvons deux périphrases désignant Eurydice : chez BALLANCHE1809, « son épouse » est précisée en apposition par « cette touchante victime [...] », alors que BALLANCHE1829 la désigne comme « celle qui veut être initié » puis par « la vierge ». L'appariement initial des syntagmes permet de choisir la relation optimale : « son épouse » et « la vierge » font partie de groupes nominaux plus larges, dont « visage » est lemme noyau, déjà en correspondance pour les deux phrases. L'algorithme prend en compte ces éléments, mais les deux périphrases restantes seront appariées entre elles, même s'il est plus correct d'intégrer l'apposition « cette touchante victime » dans la périphrase « son épouse ».

24 Différents types de graphes ont été exploités à cet effet : voir Jänicke *et al.* 2014 ; Ganaschia 2021 ; Del Lungo et Suchecka à paraître ; Suchecka, Gasiglia et Zieger 2019.

25 « [N]ul ne peut mieux que Thamyris lui-même rectifier l'histoire que tu vas dire ; nul ne peut mieux nous instruire de ce que l'on sait sur ce génie sublime, à qui les dieux inspirèrent les lois de la société, celles du bien et du beau. » (Ballanche 1827, 28.)

26 « Adieu. L'éternelle nuit m'enveloppe et m'entraîne, sans que je puisse te presser de mes mains défaillantes : hélas ! je ne suis plus à toi. » (Virgile 1897 [trad. Cabaret-Dupaty], 103).

27 « Orphée, [...] j'ai voulu te distraire de tes grandes pensées, au lieu de me borner à les embellir, et à leur donner du charme ; et voilà que les dieux m'en punissent. [...] Continue d'obéir à tes hautes destinées, que je voulais concentrer sur moi seule ; marche à ton noble but, sans te détourner ni à droite ni à gauche, et que la douleur ne fasse pas ce qu'allaient faire les condescendances de l'amour. » (Ballanche 1827, 115-116).

28 Alors que Juden (1970) accepte les interprétations découlant du monologue de la femme, il y ajoute une symbolique ésotérique : en accomplissant le rite du mariage, Eurydice s'initie sans le savoir au mystère de l'intelligence divine, mais celui-ci est trop puissant pour qu'elle puisse le supporter. En le plaçant dans le contexte plus large de l'épopée, Juden relève également une parabole de l'ascension pénible des peuples, « sortant de l'enfer de l'ignorance sociale » (148). Notre travail ne dévoile pas encore de relations permettant d'appuyer ces deux axes d'analyse. En revanche, les correspondances détectées avec la partie interprétative de BALLANCHE1809 apportent de nouvelles pistes sur le rôle de la mort d'Eurydice (voir plus loin).

29 La phrase 28 de BALLANCHE1809 (« Il y a sur la terre comme un long gémissement qui se traîne de génération en génération, depuis les premiers morts jusqu'à nous ») est appariée avec la phrase 2446 de BALLANCHE1829 (« L'homme n'a qu'une voix ; [...] et cette voix, qui n'est, le plus souvent, que celle du gémissement, passe de génération en génération, et retentit dans les siècles »). La phrase 34 (« Son existence fût demeurée incomplète ; et il aurait, au lieu de la douleur dont il fit une épreuve si cruelle, connu cette autre douleur qui n'a pas moins d'intensité, la solitude de l'âme. ») est, elle, en lien avec la 1998 (« Cette pensée continue de le suivre dans sa destinée future ; elle accompagne l'âme voyageuse [...] accomplissant ainsi, dans des existences successives, la série d'épreuves par lesquelles il faut qu'elle passe avant de pouvoir être introduite dans son état définitif »).

30 Il s'agit d'une histoire de Phœnix, racontée à Thamyris par un des prêtres d'Isis. Uranus envoie ses fils trouver l'oiseau qui est en train de construire son tombeau pour renaître de ses cendres. Pour arriver jusqu'à lui, ils doivent passer une série d'épreuves, dont celle de l'énigme : « Y a-t-il un être qui marche tantôt sur quatre pieds, tantôt sur deux, tantôt sur trois, et qui soit d'autant moins assuré dans sa démarche, qu'il semble être

mieux appuyé ? [...] » (Ballanche 1827, 253). L'extrait de la correspondance fait partie de la réponse formulée. La proximité de cette énigme, et de son contexte, et de celle posée par le Sphinx à Œdipe prouve la complexité des réseaux intertextuels observables au sein de notre corpus.

31 Pour cette visualisation, contrairement à celles des figures 6-8, la taille des nœuds (sauf celui de BALLANCHE1829 qui est mis en valeur) correspond au ratio des phrases pour lesquelles au moins une correspondance a été trouvée avec un autre texte (par rapport à la totalité des phrases d'un texte) et l'épaisseur des liens est proportionnelle au nombre de phrases en correspondance entre deux textes.

32 Bouloumié (2004) relève l'inspiration de Cocteau pour évoquer des traditions qui voient en Eurydice l'alliée involontaire des bacchantes (§ 28), alors que Kushner (1961) considère Schuré (1889) comme un point de départ des réécritures symboliques du mythe (158).

33 C'est le titre de la préface de Jean-Paul Sartre à l'*Anthologie de la nouvelle poésie nègre et malgache de langue française* (1948). À ce sujet, voir par exemple Riesz 2005 et Gyssels 2005.

34 « Autour de 1900, une nouvelle phase de gobinisme, *via* l'Allemagne, où la *Gobineau-Vereinigung* de Ludwig Schemann (1852-1938) est fondée en 1894, se développe pour des usages d'élitisme aristocratique chez Robert Dreyfus (1873-1939) ou chez le wagnérien Édouard Schuré (1841-1929), et nourrit l'anthroposociologie raciste de Vacher de Lapouge (1854-1936). » (Jarrassé 2009, § 18).

Auteurs

Karolina Suchecka

ULR 1061 ALITHILA, université de Lille, Lille, France

Karolina Suchecka est doctorante contractuelle en littérature comparée à l'université de Lille ; ses travaux qui portent sur la réécriture et l'édition intermédiaire, explorent les possibilités d'outils informatiques et du traitement automatique des langues.

ORCID [0000-0002-6485-7033](https://orcid.org/0000-0002-6485-7033)

karolina.suchecka@univ-lille.fr

Nathalie Gasiglia

UMR 8163 STL, université de Lille, Lille, France

Maîtresse de conférences habilitée à diriger des recherches à l'université de Lille, Nathalie Gasiglia est une linguiste spécialisée en métalexicographie (analyse de dictionnaires monolingues de langue générale), en linguistique de corpus et en informatique éditoriale.

ORCID [0000-0001-7149-2274](https://orcid.org/0000-0001-7149-2274)

nathalie.gasiglia@univ-lille.fr

Droits d'auteur

Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).