Annotation automatique des types de discours dans des livres audio en vue d'une oralisation par un système de synthèse

Aghilas Sini[1] Elisabeth Delais-Roussarie[2] Damien Lolive[1]
(1) Univ Rennes, CNRS, IRISA, 6 rue de Kerampont, 22300 Lannion, France
(2) UMR 6310 - Laboratoire de Linguistique de Nantes, 7 Chemin de la Censive du Tertre, 44312 Nantes, France
aghilas.sini@irisa.fr, elisabeth.delais-roussarie@univ-nantes.fr,
damien.lolive@irisa.fr

RÉSUMÉ

Pour synthétiser automatiquement et de manière expressive des livres audio, il est nécessaire de connaître le type des discours à oraliser. Ceci étant, dans un roman ou une nouvelle, les perspectives narratives et les types de discours évoluent souvent entre de la narration, du récitatif, du discours direct, du discours rapporté, voire des dialogues. Dans ce travail, nous allons présenter un outil qui a été développé à partir de l'analyse d'un corpus de livres audio (extraits de *Madame Bovary* et des *Mystères de Paris*) et qui prend comme unité de base pour l'analyse le paragraphe. Cet outil permet donc non seulement de déterminer automatiquement les types de discours (narration, discours direct, dialogue), et donc de savoir qui parle, mais également d'annoter l'extension des modifications discursives. Ce dernier point est important, notamment dans le cas d'incises de citation où le narrateur reprend la parole dans une séquence au discours direct. Dans sa forme actuelle, l'outil atteint un taux de 89 % de bonne détection.

ABSTRACT

Automatic annotation of discourse types in audio-books

To synthesize audiobooks in an expressive manner, it is necessary to know the type of discourses that have to be produced. However, in a novel or a tale, narrative perspectives and discourse types often change, moving from narrative and recitative paragraphs to direct speech, reported speech, and even dialogs. In this work, we will present a tool that was developed from the analysis of a corpus (including excerpts from *Madame Bovary* and *Les Mystères de Paris*) and that relies on paragraph as **basic unit**. It allows not only to automatically determine the type of speech (narrative speech, direct speech, dialogs), and therefore to know who is speaking, but also to annotate the extension of the discursive modifications. This later point is important, especially in the case of parentheticals with reporting verbs where the narrator speaks again in the middle of a direct speech sequence. In its current form, the tool achieves a 89 % detection rate.

MOTS-CLÉS : types de discours, discours direct, incises, annotation automatique.

KEYWORDS: discourse types, direct speech, quotation and reporting verb, automatic annotation.

1 Introduction

Pour synthétiser de façon satisfaisante et expressive des livres audio, il est important de pouvoir indiquer toute modification dans les perspectives énonciatives (les changements de locuteur dans les

séquences dialoguées, les incises de citation, etc.) par des marquages prosodiques comparables à ceux observés dans la parole naturelle (Doukhan *et al.*, 2011; Montaño *et al.*, 2013). Pour y parvenir, il est nécessaire de distinguer dans un texte les paragraphes (en entendant par paragraphe toute séquence séparée par des sauts de ligne dans le texte) selon leur type de discours, et de là d'avoir une idée précise de "qui parle". Cela conduit à classer les paragraphes selon qu'ils correspondent à des passages narratifs (1), dialogués (2), ou mixtes. De fait, dans certains paragraphes, du discours rapporté est inséré au milieu de passages narratifs (3) ou des incises de citation, correspondant à du discours narratif, apparaissent dans des discours directs, ces dernières pouvant être courtes (4a) ou relativement longues (4b). Dans ces cas mixtes, la tâche consiste à délimiter avec précision les types de discours en présence.

(1) On commença la récitation des leçons. Il les écouta de toutes ses oreilles, attentif comme au sermon, n'osant même croiser les cuisses, ni s'appuyer sur le coude, et, à deux heures, quand la cloche sonna, le maître d'études fut obligé de l'avertir, pour qu'il se mit avec nous dans les rangs. (*Madame Bovary*, chap. 10)

(2) – D'où viens-tu encore, gredin ?
– Vous êtes bien curieux, sans yeux... (*Les Mystères de Paris*, chap. 7, Partie 2)

(3) D'autre part, la mort de sa femme ne l'avait pas mal servi dans son métier, car on avait répété durant un mois : **« Ce pauvre jeune homme ! quel malheur ! »** (*Madame Bovary*, chap. 3)

(4) a. – Levez-vous, **reprit le professeur**, et dites-moi votre nom... (*Madame Bovary*, chap. 1)
b. – Débarrassez-vous donc de votre casque, **dit le professeur, qui était un homme d'esprit**. (*Madame Bovary*, chap. 1)

(5) Puis, l'ayant considéré quelques minutes d'un œil amoureux et tout humide, **elle dit vivement :** (*Madame Bovary*, chap. 18)

Alors que la détection des passages narratifs (1), des passages dialogués (2) et des discours rapportés aux milieu de passages narratifs (3) dans les paragraphes mixtes peut sembler assez triviale, du fait notamment d'indications typographiques, l'annotation des incises de citation est plus complexe, comme en témoigne une simple comparaison entre les cas (4a) et (4b). La présence d'une virgule après l'incise de citation n'est en effet pas suffisante.

Dans cet article, nous allons présenter l'outil que nous avons développé pour annoter automatiquement un texte et indiquer clairement les changements énonciatifs et discursifs. Dans un premier temps, nous présentons le corpus utilisé pour cette étude, et nous indiquons comment nous avons procédé pour délimiter les différents types de discours. Dans une seconde partie, nous fournissons des indications sur les taux d'identification obtenus et sur les problèmes résiduels.

Typologie des paragraphes	Discours direct	Discours indirect	Discours mixte
Paragraphes	1 202	844	771
Phrases	4 109	2 160	2 920
Mots	36 722	36 622	26 001
Mots orthographiquement distincts	5399	6913	4 345
Mots phonétiquement distincts	5235	6764	4 248
Syllabes	49 313	55 021	35 827
Syllabes différentes	2 692	2678	2 279
Phonèmes	111 915	124 886	80 827
Phonèmes distincts	33	33	33

TABLE 1 – Composition du corpus en fonction des types de discours

2 Corpus et procédure d'annotation des types de discours

2.1 Corpus et matériel

Pour cette étude, nous avons travaillé sur une sous-partie d'un corpus d'audiobooks comprenant 87 heures de lecture par une unique locutrice de plusieurs livres audio (Sini *et al.*, 2018). Ce dernier a été développé afin de travailler à l'amélioration de l'expressivité en synthèse par corpus et a été collecté depuis la librairie LibriVox [1]. Dans le travail présenté ici, nous avons retenu des extraits correspondants à des chapitres de deux romans français, *les Mystères de Paris* d'Eugène Sue et *Madame Bovary* de Gustave Flaubert, pour une durée totale de 10 heures. Ces extraits ont été choisis par un expert sur l'ensemble des deux œuvres. Ils ont été retenus car ils renferment de nombreux changements de perspectives discursives et énonciatives, tout en permettant d'arriver à un ensemble relativement cohérent en termes de séquences au discours direct et indirect comme indiqué dans le tableau 1.

Pour l'ensemble du corpus et de la sous-partie retenue pour cette étude, nous disposons de la transcription orthographique, de la phonétisation et de l'alignement sur le signal sonore fait automatiquement à l'aide de JTrans (Cerisara *et al.*, 2009). D'autres annotations linguistiques de nature phonologiques (découpage en syllabes, etc.) et morpho-syntaxiques (catégorisation grammaticale des mots, analyse et indication des fonctions grammaticales) sont également disponibles grâce au recours à des procédures d'annotation automatique (Candito *et al.*, 2010, 2009). L'ensemble du processus d'annotation a été mené en s'appuyant sur ROOTS (Chevelu *et al.*, 2014), ce qui permet de maintenir l'ensemble des annotations de façon cohérente.

2.2 Procédure d'annotation

L'annotation automatique des changements discursifs et énonciatifs pour un texte donné (un chapitre dans notre cas) se fait en deux phases, illustrées dans les sous-sections qui suivent :

1. https ://librivox.org

1. Classification des paragraphes en fonction des types de discours, de manière à conserver les paragraphes qui comportent des incises ;
2. Détection et délimitation des incises de citation (*dit-il*, etc.) et des amorces (*il affirma :"..."*, etc.).

2.2.1 Classification des paragraphes selon les types de discours

À partir du texte, le programme classe les paragraphes, définis sur une base typographique (passage à la ligne), en trois groupes distincts (voir fig. 1, phase 1). Il s'appuie pour cela sur des critères lexico-syntaxiques (présence de verbe de discours, etc.), ainsi que sur la ponctuation et des signes typographiques (présence de guillemets ou de tirets, etc.) :

— le groupe *Discours Direct* regroupe tous les paragraphes qui contiennent exclusivement des passages au discours direct comme dans l'exemple (2) ;
— le groupe *Narration* renferme les paragraphes ne comportant que des passages de narration ou des descriptions comme dans l'exemple (1) ;
— le groupe *Discours mixte* est composé de paragraphes qui peuvent contenir à la fois du discours direct et rapporté et du discours indirect ou de la narration. Seront présents dans ce groupe à la fois les passages narratifs dans lesquels sont insérés des discours rapportés comme dans l'exemple (3) et des passages au discours direct comprenant des incises de citation comme dans l'exemple (4).

Dans une seconde phase (voir fig. 1, phase 2), les paragraphes du groupe *Discours Mixte* sont analysés afin de déterminer les frontières exactes des changements de discours. Cette tâche est effectuée sur un mode expert par règles. Notons cependant que d'autres travaux ont eu recours à des techniques d'apprentissage automatique pour une tâche analogue (Schöch *et al.*, 2016). Cette étape va permettre d'identifier les passages au discours rapporté, souvent entre guillemets et précédés des deux points comme (3), mais aussi les incises de citation dans les passages dialogués comme (4), et les séquences amorces introduisant un passage au discours direct ou un dialogue comme (5). Parmi ces éléments, les incises de citation sont importantes car elles permettent de délimiter les changements de locuteur et de fournir des indications sur les personnages en présence et sur leur attitude.

2.2.2 Détection et annotation des incises et des amorces

À l'issue de la première phase de classification, les paragraphes *Discours mixte* sont analysés de façon détaillée pour déterminer les frontières des différents types de discours. La méthode implémentée pour détecter les incises de citation s'appuie dans un premier temps sur les travaux de (Boula de Mareüil & Maillebau, 2002), qui consiste en un ensemble d'expressions régulières. Puis on y ajoute un ensembles de règles qui visent à détecter les incises d'une manière plus détaillée, et à couvrir des cas plus complexe en s'appuyant sur l'analyse syntaxique des incises de citation décrit par (Bonami & Godard, 2008) et (Danlos *et al.*, 2010). Dans notre étude trois configurations ont été distinguées :

— les amorces de discours direct comme dans l'exemple (5) ;
— les incises de citation situées au milieu de la prise de parole d'un personnage (4a) ;
— les incises de citation placées à la fin des propos d'un personnage (4b).

S'ajoutent à ces trois configurations les cas où un discours direct est inséré dans un discours indirect

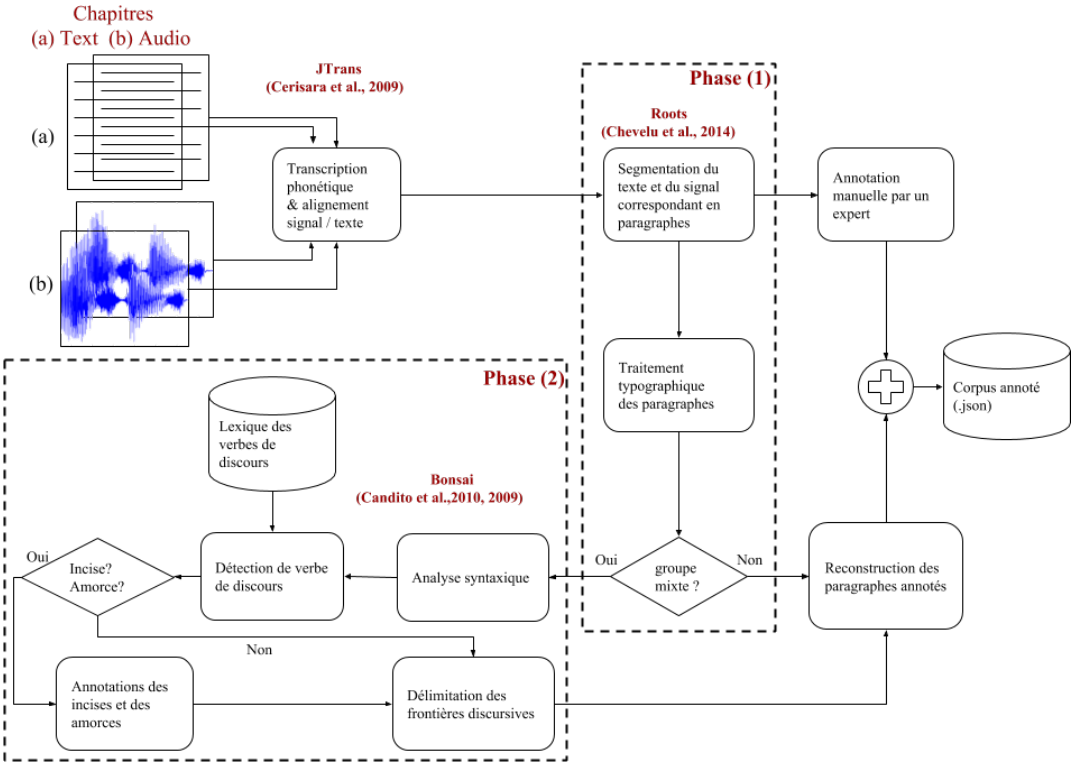

FIGURE 1 – Processus d'annotation des chapitres

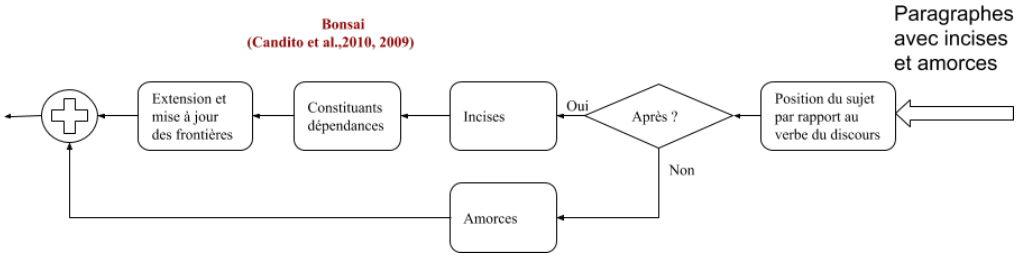

FIGURE 2 – Détection et annotation des incises et des amorces

de manière soudaine, c'est-à-dire sans amorce ou autre indication de changement de perspective discursive.

Pour analyser ces différentes configurations, il est nécessaire de regarder d'autres éléments que la seule ponctuation ou présence de tirets. Dans l'approche proposée, nous prenons en compte à la fois le résultat de l'analyseur syntaxique (Candito *et al.*, 2009) et un lexique de 327 verbes de discours (*affirmer, répéter, s'écrier, dire,* etc.). Lorsqu'un verbe de discours, généralement à la troisième personne du singulier (dans 97% des cas), est détecté, on s'intéresse à son sujet, afin de connaître sa position par rapport au verbe. Deux cas se présentent : si le sujet est à gauche du verbe (avant), il s'agit d'une amorce ; si, au contraire, il est après le verbe, on a affaire à une incise. Dans ce cas, il est important d'en établir l'extension, les incises pouvant être courtes (comme (4a)) ou relativement longue (voir (4b)). Pour ce faire, on s'appuie sur la ponctuation, mais également sur l'analyse syntaxique, notamment pour les éléments à droite du verbe qui peuvent dépendre du sujet comme dans le cas d'une apposition ou d'une relative (voir exemple (4b)). Le processus complet est illustré dans la fig. 2.

3 Résultats et analyse

Les résultats obtenus par cet algorithme de détection des types de discours sont donnés dans le tableau 2. Les performances ont été estimées avec trois mesures : la précision, le rappel, et la F-mesure.

L'algorithme permet une bonne classification des paragraphes (92,19 % de bonne détection ou F-mesure) à l'issue de la phase 1. Pour l'analyse des incises (801 annotées manuellement sur le corpus), les performances ont été calculées en distinguant deux niveaux d'annotation. La détection simplifiée - laquelle repose sur la prise en compte des verbes de discours et de la ponctuation (voir fig.1, phase 2) - ne permet pas de délimiter avec précision l'extension des incises (F-mesure : 86,3 %), des erreurs apparaissant lorsqu'une relative ou une apposition dépendent du sujet. La prise en compte de l'analyse syntaxique et des dépendances comme indiqué dans la figure 2 permet, en revanche, d'affiner les résultats et de les améliorer, si bien qu'on atteint, à l'issue de cette détection précise, un score de 89,09 % (ce dernier comprenant le type de discours et l'extension des incises).

	Précision	Rappel	F-mesure
Annotations des paragraphes (Phase 1)	92,6	91,2	92,19
Détection simplifiée des types de discours (direct, indirect, incises et amorces)	87,5	85,2	86,33
Détection précise des incises (avec délimitation fine)	89,7	88,5	89,09

TABLE 2 – Résultats de la détection et de l'annotation des changements discursifs

Une étude des erreurs permet d'isoler deux cas :

— ceux où le verbe de discours prend la forme d'un participe, et non d'un verbe conjugué, comme dans l'exemple (6). L'algorithme a en effet tendance à limiter l'incise à *arrêta*.

(6) – Cinq cents vers à toute la classe ! **exclamé d'une voix furieuse, arrêta, comme le Quos ego, une bourrasque nouvelle.** (*Madame Bovary, Chapitre 1*)

— ceux où l'analyse syntaxique effectuée est erronée comme dans l'extrait (7). La complexité et les enchâssements syntaxiques dans l'incise rendent son analyse difficile.

(7) – Oui... j'entends bien ; vous voulez que je vous mène à sa porte... et puis à son lit... et puis que je vous dise où frapper, et puis que je vous guide le bras, n'est-ce pas ? Vous voulez enfin me faire servir de manche à votre couteau !... vieux monstre ! **reprit Tortillard avec une expression de mépris, de colère et d'horreur qui, pour la première fois de la journée, rendit sérieuse sa figure de fouine, jusqu'alors railleuse et effrontée.** On me tuerait plutôt... entendez-vous... que de me forcer à vous conduire chez votre femme. (*Les Mystères de Paris,Chapitre 7, Partie 2*)

Les résultats obtenus sont moins bons que ceux présentés dans des travaux de classification des discours direct et indirect reposant sur des algorithmes d'apprentissage automatique (voir (Schöch *et al.*, 2016) qui obtient une F-mesure de 93.9 % en utilisant l'algorithme "Forêt d'arbres décisionnels"). Ceci étant, cette différence est à prendre avec précaution car les objectifs recherchés ne sont pas exactement les mêmes. La procédure développée par (Schöch *et al.*, 2016) vise à dire si chaque phrase est au discours direct ou indirect, mais n'isole pas les incises de citation, les amorces ou les passages au discours direct dans une séquence narrative. Cela s'explique par une différence fondamentale d'objectifs : alors que (Schöch *et al.*, 2016) veut classifier les œuvres sur des bases littéraires en s'appuyant sur la présence ou non de discours direct, nous souhaitons précisément savoir *qui parle* et à quel moment précis s'opère le changement. De plus, les différences de résultats peuvent s'expliquer par la méthode retenue : alors que (Schöch *et al.*, 2016) prend comme unité de base la phrase, nous prenons le paragraphe dans le but d'indiquer tout changement discursif dans un même paragraphe. En outre, nous avons recours à un modèle expert avec l'usage de règles, tandis qu'ils utilisent des procédures d'apprentissage automatique. Il serait d'ailleurs intéressant d'utiliser des procédures analogues à celles retenues par (Schöch *et al.*, 2016), mais en gardant les mêmes objectifs, à savoir déterminer précisément où s'opèrent les changements discursifs.

4 Conclusion et perspectives

Dans cet article, nous avons proposé un algorithme qui permet d'annoter automatiquement et avec précision les changements discursifs dans les livres audio. Les performances de l'outil sont relativement encourageantes, mais des erreurs subsistent dans les cas syntaxiquement complexes. Nous envisageons de nous appuyer sur le signal audio pour avoir une meilleure appréhension des changements discursifs en général, et pour mieux délimiter l'extension des incises dans les cas complexes. Cela pourrait ensuite servir de base pour une classification à l'aide d'outils d'apprentissage automatique.

5 Remerciements

Le travail présenté ici a été soutenu par l'opération PPC 7 du Labex "Empirical Foundations in Linguistics" (ANR-10-LABX-0083). Il a également bénéficié du soutien financier de l'Agence Nationale de la Recherche dans le cadre du projet ANR SynPaFlex (ANR-15-CE23-0015).

Références

BONAMI O. & GODARD D. (2008). Syntaxe des incises de citation. In *Actes du premier Congrès Mondial de Linguistique Française*, p. 2395–2408, France.

BOULA DE MAREÜIL P. & MAILLEBAU E. (2002). Traitement des incises en français : capture automatique et modèle prosodique. In *XXIVèmes Journées d'Étude sur la Parole, Nancy*.

CANDITO M., CRABBÉ B., DENIS P. & GUÉRIN F. (2009). Analyse syntaxique du français : des constituants aux dépendances. In *16e Conférence sur le Traitement Automatique des Langues Naturelles - TALN 2009*, Senlis, France.

CANDITO M., NIVRE J., DENIS P. & ANGUIANO E. H. (2010). Benchmarking of statistical dependency parsers for french. In *Proceedings of the 23rd International Conference on Computational Linguistics : Posters*, p. 108–116 : Association for Computational Linguistics.

CERISARA C., MELLA O. & FOHR D. (2009). Jtrans, an open-source software for semi-automatic text-to-speech alignment. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association-Interspeech 2009*.

CHEVELU J., LECORVÉ G. & LOLIVE D. (2014). ROOTS : a toolkit for easy, fast and consistent processing of large sequential annotated data collections. In *Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland.

DANLOS L., SAGOT B. & STERN R. (2010). Analyse discursive des incises de citation. In *2ème Congrès Mondial de Linguistique Française - CMLF 2010*, La Nouvelle Orléans, United States : Institut de Linguistique Française.

DOUKHAN D., RILLIARD A., ROSSET S., ADDA-DECKER M. & D'ALESSANDRO C. (2011). Prosodic Analysis of a Corpus of Tales. p. 3129–3132, Florence, Italy : International Speech Communication Association (ISCA).

MONTAÑO R., ALÍAS F. & FERRER J. (2013). Prosodic analysis of storytelling discourse modes and narrative situations oriented to text-to-speech synthesis. In *Eighth ISCA Workshop on Speech Synthesis*.

SCHÖCH C., SCHLÖR D., POPP S., BRUNNER A., HENNY U. & TELLO J. C. (2016). Straight talk ! automatic recognition of direct speech in nineteenth-century french novels. In *Digital Humanities 2016 : Conference Abstracts*, p. 346–353.

SINI A., LOLIVE D., VIDAL G., TAHON M. & ÉLISABETH DELAIS-ROUSSARIE (2018). Synpaflex-corpus : An expressive french audiobooks corpus dedicated to expressive speech synthesis. In *Language Resources and Evaluation Conference (LREC) to appear*.