Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

'Il était une fois' les patterns prosodiques des contes de fée

Rim Abrougui[2], Katarina Bartkova[1,2]
(1) Atilf UL, 54000 Nancy, France
(2) Université de Lorraine, 54000 Nancy, France
{rim.abrougui;katarina.bartkova}
@univ-lorraine.fr

RÉSUMÉ

Nous étudions ici la différence des patterns prosodiques entre deux styles de lecture, un que nous appelons 'lecture littéraire neutre' et un style de 'lecture des contes'. Les données appartenant au style de 'lecture de contes' comportent deux sous-ensembles, des contes destinés aux jeunes enfants (0-6 ans) et des contes destinés aux enfants plus âgés et aux adultes. Les corpus ont été manuellement annotés avec des étiquettes sémantico-prosodiques exprimant des attitudes, des émotions et d'autres styles prosodiques. Une analyse détaillée des caractéristiques prosodiques nous a permis d'identifier les traits pertinents des patterns intonatifs des différentes étiquettes et des différents styles de lecture. Une quantification vectorielle, utilisant essentiellement des informations de F0, a été utilisée pour dégager les patterns prosodiques typiques correspondant aux différentes étiquettes. Une classification automatique basée sur des paramètres prosodiques a montré une bonne identification des étiquettes quand leur fréquence était suffisamment élevée pour obtenir une modélisation robuste.

ABSTRACT

'Once upon a time' prosodic patterns of fairy tales
Here, we study the difference in prosodic patterns between two reading styles, one that we call "neutral literary reading" and one corresponding to "storytelling reading". The data concerned with the "storytelling" style include two subsets, tales intended for young children (0-6 years) and tales intended for older children and adults. The corpora have been manually annotated with semantic-prosodic labels expressing attitudes, emotions and other prosodic styles. A detailed analysis of the prosodic characteristics allowed to identify the relevant features of the intonation patterns associated to the different labels and different reading styles. A vector quantization procedure, essentially using F0 values information, was used to identify the typical prosodic patterns of the different semantic-prosodic labels. Using a tree classifier, an automatic classification based on the prosodic parameters showed a good identification of the labels when their frequency was high enough to obtain a reliable modeling.

MOTS-CLÉS : style de parole, prosodie, étiquette sémantico-prosodique
KEYWORDS: speech style, prosody, prosodico-semantic label

Volume 1 : Journées d'Études sur la Parole, pages 1–9.
HAL : hal-02798508.

1 Introduction

Le style de voix peut être identifié en grande partie par la prosodie utilisée. Si l'étude des styles est souvent réalisée sur l'écrit, elle est aussi omniprésente dans l'analyse orale. Les variations phoniques dues aux situations ou aux individus constituent les objets primordiaux dans le domaine de la phonostylistique. À travers la variation codée des paramètres prosodiques, se réalisent les différents styles de voix reflétant des informations implicites dans la communication qui restent souvent inaccessibles dans le sens explicite d'un énoncé (Ackerman, 1981). La variation de ces paramètres contribue à la perception naturelle de la parole car elle permet d'identifier les attitudes et les émotions exprimées par le locuteur. L'analyse et la modélisation de ces paramètres demeurent un défi permanent pour les chercheurs afin d'introduire le naturel dans la synthèse automatique de la parole (Gelin et al., 2010, Doukhan, 2007, Cambell, & Mokhtari, 2003).

L'analyse acoustique des styles de voix a démontré que les paramètres prosodiques varient systématiquement en fonction de la situation de communication et que ce sont ces variations qui reflètent l'expressivité (Beller, 2009). Cependant l'expressivité couvre un domaine encore plus vaste puisqu'elle peut se définir comme un indicateur vocal d'un état émotionnel, d'un style de parole ou même d'une intention implicite (Granström, & House, 2005). Ainsi, certains styles partagent les mêmes comportements prosodiques, comme par exemple la ressemblance dans l'utilisation des patterns de F0 entre les émotions 'tristesse' et 'dégout' (Scherer, 2005, Bartkova et al., 2016). Cette complexité des paramètres peut expliquer la multiplication des études sur la prosodie en relation avec l'expressivité abordée souvent dans un but de modélisation de l'expressivité pour les technologies vocales (Chella et al., 2008, Burkhardt, 2011).

Dans notre étude, nous avons choisi de nous intéresser à la narration des contes de fée. Cela nous a apparu pertinent pour l'étude des styles de voix car le conte, qui est par définition un récit narratif rattaché primordialement à l'écrit, est tout d'abord oral par tradition, ce que témoigne le style de narration des conteurs. En effet, lorsqu'un conteur prend la parole pour lire une histoire, il exploite la variation de sa voix, et celle de sa prosodie, pour attirer l'attention de son public. De nombreuses études se sont intéressées au style des contes souvent dans le but de leur utilisation dans les technologies de la parole (Doukhan et al., 2011 ; Sarkar et al., 2014) ou pour analyser d'une façon détaillée les phénomènes prosodiques pertinents pour l'expression de ce style de parole (Delais-Roussarie & Yoo, 2014).

L'autre style de parole étudié est la lecture de texte littéraire que l'on peut considérer comme un style plus neutre. La lecture 'neutre' se définit comme la parole préparée, elle est donc construite, et tend vers l'écrit (Bazillon et al., 2008). En cela, la lecture 'neutre' est comme la lecture des contes puisque les deux sont des genres narratifs. Elle est définie en tant que 'neutre', car son expressivité demeure à un degré qui avoisine '0', (Tao, 2006, Inanoglu, 2009). C'est pour cela que nous n'enregistrons pas de variations importantes de ses patterns prosodiques. Nous avons analysé la lecture 'neutre' pour la comparer avec le style de lecture des contes en mettant en évidence les points communs et les points de différence entre ces deux styles.

Le but de notre étude est de comparer les patterns prosodiques des contes destinés pour les auditeurs jeunes (narration visant un public d'enfants de moins de 6 ans) avec les contes destinés à des auditeurs plus âgés et des adultes. Notre objectif est d'étudier si une différence de style existe et si elle existe, nous voudrions vérifier si elle est portée par la majorité des patterns intonatifs ou si c'est la fréquence d'occurrence de certains patterns qui caractériserait un style de voix. La deuxième

question que nous nous posons dans cette étude, concerne la possibilité d'utilisation de modèles de patterns intonatifs pour annoter des corpus de parole. En effet, il n'est pas facile pour un annotateur humain de faire abstraction de contenu sémantique et de se focaliser exclusivement sur les caractéristiques prosodiques de la parole. Or, cela serait tout à fait envisageable si une approche automatique était mise en place. Mais pour une telle approche une bonne identification automatique des patterns prosodiques s'avère indispensable.

2 Corpus de parole

Notre corpus est constitué de 3 styles de lecture : style conte de fée destiné aux jeunes auditeurs ('CP', pour contes pour les petits), style conte de fée destiné aux auditeurs plus âgés ('CG', pour contes pour les grands) et style de lecture que nous considérerons comme un style neutre de lecture de texte littéraire ('LN'), destiné aux adultes et adolescents. La quasi-totalité de nos données sonores ont été obtenues à partir du site de la littérature audio (http://www.litteratureaudio.com/).

2.1 Traitement prosodique du corpus

Des traitements manuels, semi-automatiques et automatiques ont été appliqués sur les données. Les données de parole ont été transcrites orthographiquement (avec l'outil Transcriber) et ont été segmentées automatiquement par le logiciel Astali (http://ortolang108.inist.fr/astali/). La segmentation automatique a été manuellement vérifiée et corrigée quand cela semblait nécessaire. Des paramètres de F0 et d'énergie ont été calculés avec le logiciel Aurora (ETSI, 2005). À partir de ces données, différentes caractéristiques prosodiques (pentes et niveau de F0, durée et énergie vocalique normalisées …) ont été calculées. Ces données prosodiques ont été utilisées pour segmenter automatiquement la base de données en groupes intonatifs (GI) en évaluant des marques prosodiques sur les syllabes finales des unités lexicales (Bartkova et al., 2012). Pour cette segmentation une vérification manuelle a été également réalisée.

Les GI ont été utilisés comme fenêtre d'observation des paramètres prosodiques devenant également des unités d'étiquetage. Chaque GI a été étiqueté (voir section 2.2), et a été représenté par 6 valeurs de F0, 3 valeurs correspondant aux valeurs de F0 sur la dernière voyelle (début, milieu et fin), une valeur correspondant à la première voyelle et la valeur la plus basse et la valeur la plus haute parmi les valeurs restantes, tout en respectant l'ordre temporel de ces valeurs. Si le groupe prosodique ne contenait pas assez de voyelles (4 au minimum pour obtenir les 6 valeurs de F0), alors une interpolation a été effectuée à partir des valeurs disponibles. La décision de représenter la dernière voyelle par 3 valeurs de F0 a été prise dans un souci de représenter correctement un pattern intonatif de forme complexe (circonflexe, concave, etc.) présent surtout dans le style des contes.

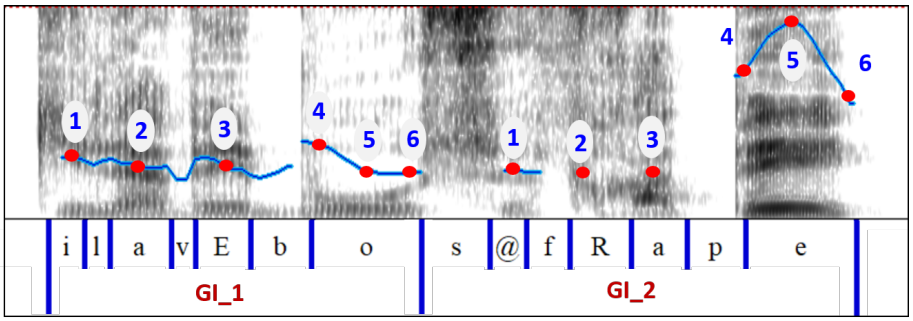

Figure 1 : Exemples de formes intonatives complexes
et des (6) valeurs de F0 retenues pour chaque GI

Le corpus CP comporte 6 contes lus par 4 locuteurs. Le corpus CG comporte 3 contes lus par 3 locuteurs et le corpus LN comporte 7 enregistrements lus par 6 locuteurs. Dans chaque corpus, les locuteurs-conteurs était féminins et masculins.

	Nb Mots	VA syl/sec	Durée (ms) V[-acc]	Durée (ms) V[+acc]	GS (nb syll)	GI (nb syll)	Pause (ms)	Étendue voc. (st)
CP	4605	4.6	79 (±33)	155 (±61)	4	2.4	627	24
CG	8379	4.9	75 (±32)	147 (±62)	5	2.6	517	26
LN	10995	5.5	68 (±23)	131 (±40)	6	3.0	639	16

Table 1 : Caractéristiques des corpus utilisés (voir explications dans le texte)

La table 1 indique la taille des corpus, ainsi que quelques informations sur l'organisation temporelle et les caractéristiques de la fréquence fondamentale de nos trois corpus. Le style le plus rapide (aussi bien par la vitesse d'articulation (VA) que par la durée des voyelles non-accentuées (V[-acc]) et accentuées (V[+acc]) est le style LN, et le plus lent est le style CP. C'est également le style CP qui a les groupes de souffles (GS) et les groupes intonatifs (GI) les plus courts (constitués en moyenne de 4 et de 2.4 syllabes respectivement). Le paramètre de l'étendue vocale du locuteur (en semi-tons dans le tableau) s'avère être très pertinent pour différencier les trois styles de parole. En effet elle est nettement plus large pour les styles de narration de contes (CP et CG) que pour le style LN. L'étendue vocale, pour un locuteur, correspond ici à la différence entre la valeur de F0 la plus élevée et la valeur la plus basse. Afin de pallier les éventuelles erreurs de détection de F0, les 2% des valeurs les plus basses et les plus élevées de F0 ont été écartées de l'estimation de l'étendue vocale.

2.2 Étiquetage sémantico-prosodique des corpus

Nos corpus ont été étiquetés manuellement par des étiquettes sémantico-prosodiques reflétant différentes attitudes et émotions ainsi que des réalisations prosodiques liées à un style donné (Akpossan-Confiac, 2007). L'étiquetage a été réalisé par 2 étiqueteurs experts qui ont attribué à chaque GI une étiquette. Les étiquettes ont été structurées sur trois niveaux (voir la figure 2) nous permettant d'introduire un lissage en cas d'étiquettes peu fréquentes dans les corpus. Les étiquettes sémantico-prosodiques représentaient des attitudes et des émotions positives et négatives (Vaudable, 2012) et des réalisations prosodiques plus neutres.

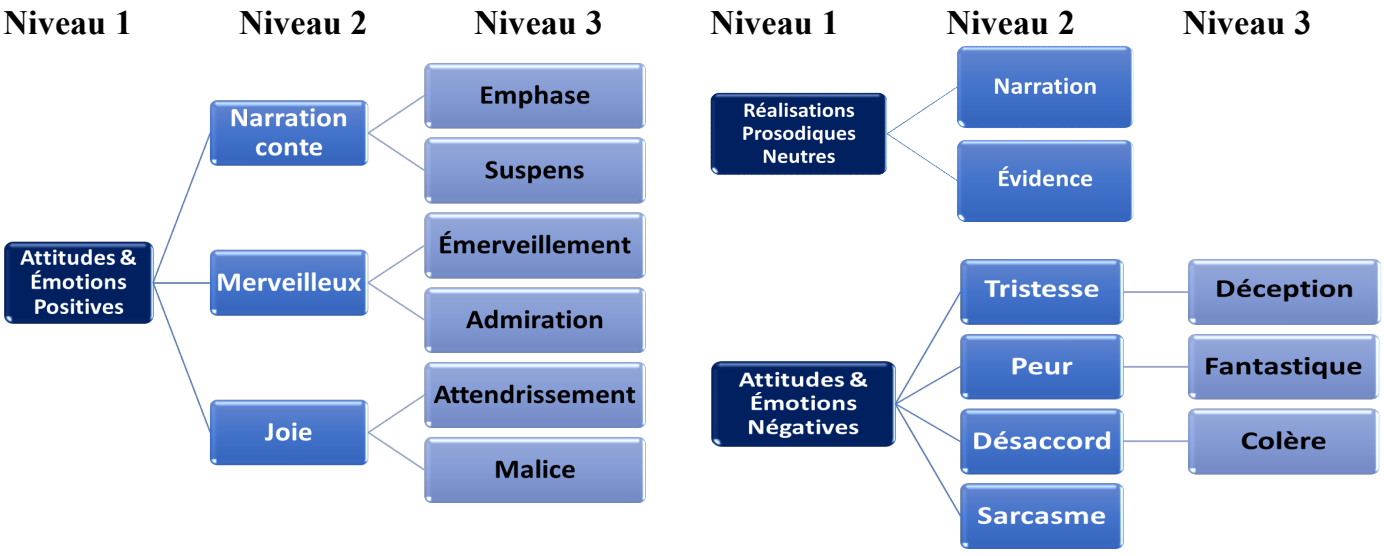

Figure 2 : Etiquettes sémantico-prosodiques

Lors de l'étiquettage sémantico-prosodique, l'attention des étiqueteurs devait se focaliser avant tout sur le pattern prosodique et non sur le contenu sémantique de l'énoncé. Les étiquettes sont distribuées d'une façon inégale dans les différents corpus. Dans le style LN, la très grande majorité (70%) des étiquettes correspondait au style 'narration'. Le nombre d'étiquettes de 'narration' avoisinait à peu près 50% dans les deux styles de contes (CP & CG). Pour les étiquettes restantes, la distribution était peu équilibrée (voir la Figure 3), les fréquences d'occurrences les plus élevées ont été observées pour l'étiquette '**suspense'** dans le style CG et '**évidence'** dans le style CP.

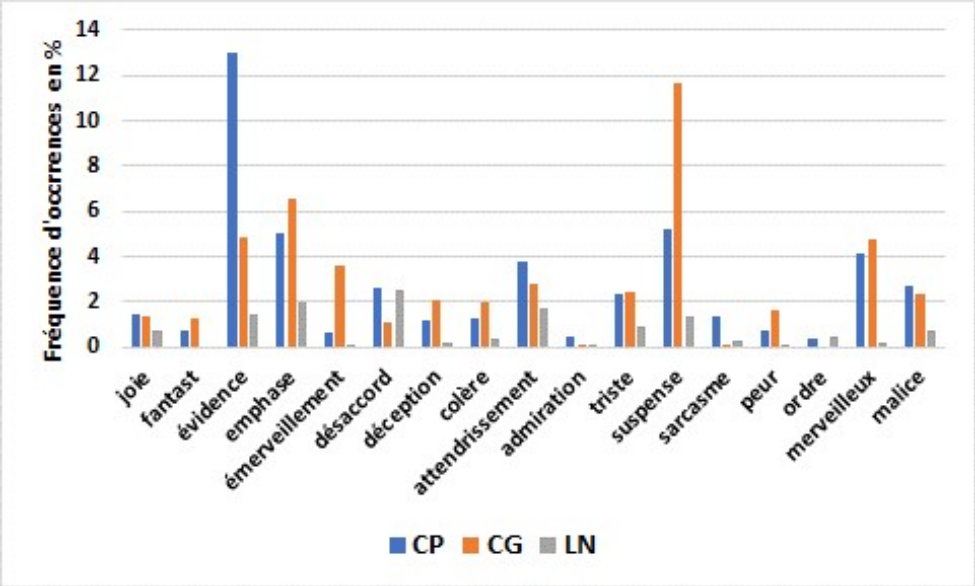

Figure 3 : Fréquence d'occurence des étiquettes sémantico-prosodiques
(l'étiquette « narration », n'est pas représentée sur la figure)

Pour évaluer le niveau de l'accord inter-annotateur, un test sur un sous-ensemble réduit de données représentant chaque style a été mené avec 6 étiqueteurs non entrainés (étudiants en linguistique du niveau master). Le test du Kappa a été réalisé à partir de ces différentes annotations et les résultats ont montré un accord inter annotateur faible (0.3) pour l'ensemble des étiquettes posées. Pour les 3 étiquettes ('*emphase', 'désaccord' et 'narration'*) l'accord Kappa était modéré (>0.4). Ce test montre un accord d'étiquetage acceptable pour un groupe non-entrainé mais spécialiste du langage et laisse penser qu'un accord plus élevé existe entre des annotateurs entraînés à la tâche ce qui était le cas lors de l'étiquetage de nos corpus.

3 Patterns intonatifs

Afin d'identifier les patterns intonatifs typiques correspondant à nos étiquettes sémantico-prosodiques, nous avons utilisé la technique de la quantification vectorielle pour classer tous les patterns intonatifs associés aux étiquettes. Avant d'appliquer la quantification vectorielle, nous avons quantifié les valeurs de F0. Pour cela nous avons estimé une étendue vocale théorique des locuteurs, comme étant supérieure d'une octave et inférieure d'une demi-octave aux valeurs médianes de F0 de chaque locuteur (De Looze, Hirst, 2014). L'étendue vocale théorique a été divisée en 10 niveaux permettant d'exprimer les valeurs de F0 par un niveau tonal de 1 à 10 à l'intérieur de l'étendue vocale théorique. Lorsqu'une valeur de F0 sort de l'étendue vocale théorique, sa valeur quantifiée (son niveau tonal) sera supérieure à 10 ou inférieure à 1.

3.1 Patterns intonatifs représentatifs

Pour représenter chaque étiquette sémantico-prosodique par son pattern intonatif, nous avons sélectionné le centroïde de la classe obtenue par la quantification vectorielle qui regroupait un grand nombre d'éléments et qui présentait un écart-type faible, c'est-à-dire, dont la variation des paramètres des éléments regroupés était faible. Ainsi nous avons obtenu un pattern typique, représentatif par étiquette.

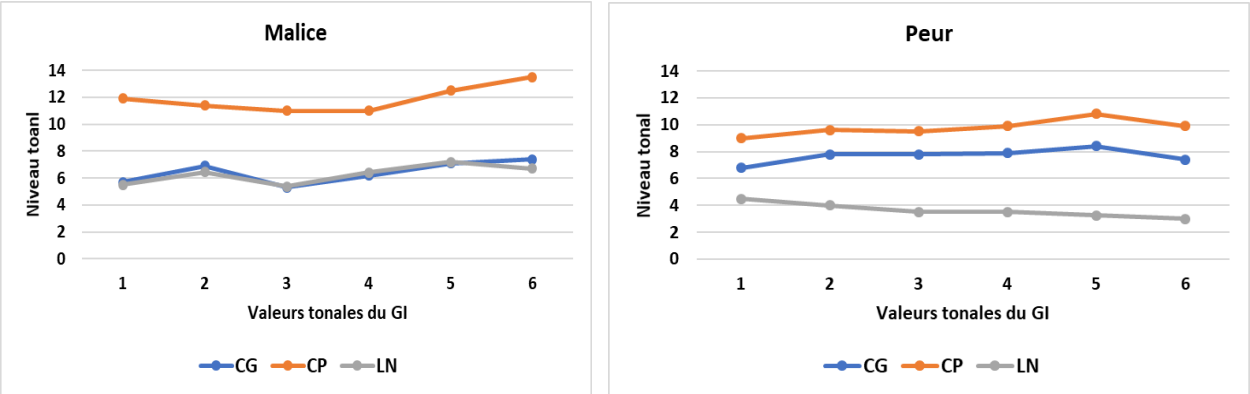

Figure 4 : Exemple de patterns intonatifs pour deux étiquettes '*Peur*' et '*Malice*' et pour les 3 styles (1 Niveau tonal = 1,8 semi-tons)

La figure 4 présente les patterns intonatifs typiques obtenus avec la quantification vectorielle pour deux étiquettes sémantico-prosodiques. Ces exemples montrent une similitude dans l'expression de la '*peur*' dans le style des contes (les deux patterns contiennent des valeurs de F0 d'un niveau élevé – valeurs comprises entre le niveau 7 et 11) alors que dans le style LN l'expression de la '*peur*' est plus 'contenue', exprimée sur un niveau tonal plus bas (entre les niveaux 3 et 4.5). Quant à la '*malice*', cette attitude est exprimée avec le même pattern intonatif dans les styles CG et LN et sur un niveau très élevé (plus haut que l'étendue vocale estimée, entre 11 et 13,5) pour le style CP.

Émotions & attitudes positives									
	Narr.	**Malice**	**Merv.**	**Admir.**	**Attendr.**	**Emphase**	**Évidence**	**Joie**	**Susp.**
CP	**2 ,7**	**11,9**	1,4	**7,3**	**1,4**	**7,0**	**3,1**	**8,5**	**1,8**
CG	4,9	6,4	1,2	5,7	1,4	5,6	1,8	6,9	7,4
LN	3,6	6,3	2,5	4,2	3,0	4,3	2,8	6,9	4,6

Émotions & attitudes négatives					
	Peur	**Tristesse**	**Colère**	**Déception**	**Désaccord**
CP	**9,8**	1,9	**7,3**	**2,1**	**8,0**
CG	7,7	1,0	6,6	5,2	2,0
LN	3,6	2,7	4,4	3,1	5,0

Table 2 : Niveau tonal moyen des étiquettes sémantico-prosodiques
(3,3 correspond à la valeur F0 médiane du locuteur)

La variation des niveaux de F0 à l'intérieur d'un même pattern (sur un même GI), n'était pas très importante – les valeurs variaient modérément d'une voyelle à l'autre, par conséquent, il est permis de considérer que les F0 des patterns peuvent être caractérisés par leur niveau intonatif moyen. Pour pouvoir comparer d'une façon plus synthétique les patterns intonatifs correspondant aux différentes étiquettes des trois corpus étudiés, nous les représentons dans la table 2, par leur valeur moyenne,

calculée sur les patterns représentatifs. Uniquement les étiquettes qui ont été observées dans les trois corpus, sont indiquées dans le tableau et seront analysées ci-après.

Les valeurs des tonalités moyennes des différents patterns représentatifs nous montrent que les patterns du style CP ont le niveau tonal le plus extrême (le plus haut – couleur rouge dans la table 2 ou le plus bas – couleur bleue dans la table 2) dans presque 86% des cas. Quant au style LN, le niveau tonal de ses patterns intonatifs se situe majoritairement (dans 80% des cas) sur un niveau bas. Le style CG reste dans la plupart des cas entre les niveaux tonals de ces deux styles, pouvant être considéré comme un style 'transitoire' entre le style CP et le style LN.

Si conformément à notre arborescence un regroupement plus global des étiquettes au niveau 1 est effectué (émotions & attitudes positives versus émotions & attitudes négatives) alors nous obtenons pour les patterns intonatifs positifs un niveau tonal moyen de 10.5 pour le style CP, un niveau de 6,2 pour le style CG et finalement un niveau tonal moyen de 4.4 pour le style LN. Pour les patterns intonatifs négatifs, nous obtenons de nouveau le niveau tonal le plus haut, 7,9, pour le style CP, le niveau tonal le plus bas, 2.3, pour le style LN, et de nouveau un niveau tonal intermédiaire de 6,7 pour le style CG.

3.2 Classification automatique

L'étiquetage sémantico-prosodique des données de parole est fastidieux et comme indiqué dans la section 2.2, l'accord inter-annotateur pour les annotateurs non-experts n'est pas très élevé car la tâche reste très complexe. Par conséquent, l'utilisation d'une annotation automatique, basée essentiellement sur les indices prosodiques, serait très utile. Pour tester la faisabilité et la fiabilité d'une telle approche, nous avons réalisé une classification automatique de nos étiquettes afin d'évaluer si une identification automatique des étiquettes à partir des patterns intonatifs et de la durée quantifiée de la voyelle accentuée (voyelle finale) était envisageable. Pour la classification automatique, nous avons utilisé l'arbre de décision J48, disponible dans le toolkit weka (https://sourceforge.net/projects/weka/). Pour les tests, nous avons, d'une part, regroupé toutes les étiquettes de nos différents corpus ('Tous corpus'), et d'autre part, nous avons également classifié les étiquettes de chacun des trois corpus séparément. Comme déjà mentionné, notre corpus LN est déséquilibré car la majorité des étiquettes de ce corpus correspond à la '*narration'*. Pour pallier au moins en partie ce biais, nous n'avons gardé qu'un cinquième de ces étiquettes dans les tests de classification. Les tests ont été réalisés pour les étiquettes se trouvant au niveau 3 (feuilles) et au niveau 2 (nœuds immédiatement supérieurs) de notre arborescence d'étiquettes (cf. figure 2).

	Tous corpus	**CP**	**CG**	**LN**
Niveau 3	45%	27%	33%	49%
Niveau 2	50%	43%	38%	55%

Table 3 : Taux d'identification correcte des étiquettes sémantico-prosodiques

Les faibles résultats du 'Niveau 3' sont imputables au nombre réduit d'exemples pour certaines étiquettes, sauf l'étiquette '*narration'*. Quand un lissage est opéré sur les étiquettes ('Niveau 2') on obtient des résultats légèrement (LN) ou significativement (CP, CG) meilleurs pour les différents corpus

Il nous a semblé intéressant de tester si le pattern intonatif de l'étiquette 'narration', très fortement présente dans nos corpus, est identifiable comme appartenant à un style de parole particulier, c'est-à-dire pouvant être identifié comme '*narration-CP'*, '*narration-CG'* ou '*narration-LN'*. Nous avons

effectué un test de détection uniquement sur cette étiquette en sélectionnant un sous-ensemble du corpus LN pour éviter le biais de la surreprésentation des étiquettes appartenant à ce corpus. L'identification de l'appartenance du pattern de *narration* à un corpus particulier a atteint 65%. Les résultats plus détaillés sont dans le tableau de confusion de la table 4.

	CP	**CG**	**LN**
CP	**40%**	**44%**	16%
CG	22%	**58%**	20%
LN	6%	14%	**80%**

Table 4 : Classification de l'étiquette '*narration'*

Comme cela apparait dans la table 4, les étiquettes '*narration'* sont fortement dépendantes des corpus. Les étiquettes '*narration-CP'* ont été souvent confondues avec les étiquettes '*narration-CG'*. Les deux corpus étant des contes, leurs différences prosodiques sont probablement exprimées par d'autres étiquettes que celle de '*narration'*. Quand des confusions existent dans l'identification de l'étiquette '*narration-CG',* elles se font aussi bien avec les étiquettes '*narration-CP'* qu'avec les étiquettes *'narration-LN'*, suggérant ainsi un positionnement intermédiaire de ces étiquettes entre ces deux styles, CP et LN. Finalement, l'étiquette '*narration-LN'* est très bien identifiée comme appartement au style de lecture et uniquement une très faible confusion (6%) existe avec les étiquettes *'narration-CP'* et une confusion très légèrement plus élevée avec les étiquettes *'narration-CG'* (14%).

4 Conclusion

Le but de cette étude était d'étudier 3 styles de parole lue, deux appartenant à la lecture de contes et un au style de lecture neutre de textes littéraires. Le corpus de contes contenait des contes destinés aux enfants jeunes et des contes destinés aux enfants moins jeunes et aux adultes. Une différence des patterns intonatifs a été observée parmi ces 3 styles, le style CP étant le plus extrême et le style LN est le plus neutre. Il apparaît de cette étude, que les styles de lecture des contes et de lecture neutre n'ont pas de socle stylistico-prosodique commun important, ainsi, par exemple, même pour l'étiquette '*narration',* très fréquente dans chaque style, les caractéristiques prosodiques sont fortement dépendantes de chaque style. Par conséquent, le style de parole doit être considéré comme un phénomène plus global que simplement l'apparition plus au moins fréquente d'étiquettes exprimant des attitudes ou des émotions.

Nous voulions également tester si l'identification automatique des patterns prosodiques représentant des étiquettes sémantico-prosodiques était suffisamment fiable, afin de les utiliser dans l'avenir pour annoter des corpus de différents styles de voix. En effet, une telle annotation bien que nécessaire, reste chronophage et fastidieuse. De plus, un annotateur humain a souvent du mal à faire abstraction du contenu sémantique de la parole et à se focaliser uniquement sur les patterns intonatifs. Cependant, pour développer une approche d'étiquetage automatique performante, il faudrait disposer de corpus plus larges et représentatifs des différentes étiquettes ; en effet, le nombre limité d'occurrences de certaines étiquettes dans nos corpus n'a pas permis d'obtenir des modèles prosodiques suffisamment fiables pour une classification robuste.

Références

ACKERMAN B.P., (1981). Young Children's Understanding of a Speaker's Intentional Use of a False Utterance, *Developmental Psychology, 17, (pp. 472-480)*

AKPOSSAN-CONFIAC, J., & DELUMEAU, F. (2007). Comment la prosodie donne du sens aux interjections?, *Interfaces discours-prosodie : actes du 2ème Symposium international IDP07 & Colloque Charles Bally, Université de Genève,* (pp. 335-347)

BARTKOVA K., DELAIS-ROUSSARIE E., & SANTIAGO VARGAS F., (2012). PROSOTRAN : Un Système d'annotation Symbolique Des Faits Prosodiques Pour Les Données Non-Standards. *In Proceedings of the Joint Conference JEP-TALN-RECITAL*, *Volume 1: JEP, Grenoble, France: ATALA/AFCP*, *(pp. 601– 608)* https://www.aclweb.org/anthology/F12-1076

BARTKOVA K., JOUVET D., DELAIS-ROUSSARIE E., (2016) Prosodic Parameters and Prosodic Structures of French Emotional Data , 8th *Speech Prosody*, Boston, United States

BAZILLON T., JOUSSE V., BÉCHET F., ESTÈVE Y., LINARÈS G. et LUZZATI D., (2008) La parole spontanée : transcription et traitement, *In Revue Traitement Automatique des Langues (TAL), volume 49, (pp. 47–67)*

BELLER, G. (2009). *Analyse et modèle génératif de l'expressivité. Application à la parole et à l'interprétation musicale*, Thèse de doctorat, Université Paris VI, IRCAM.

BURKHARDT F. (2011). An Affective Spoken Storyteller, In *In Proceedings Interspeech*, (*pp. 3305-3306) https://www.isca-speech.org/archive/interspeech_2011/i11_3305.html*

CAMPBELL, N., & MOKHTARI, P. (2003). Voice Quality: the 4th Prosodic Dimension, *15th ICPhS, (pp. 2417-2420)*

CHELLA A., BARONE R.E., PILATO G., SORBELLO R. (2008). An Emotional Storyteller Robot. *In AAAI Spring Symposium on Emotion, Personality and Social Behavior*, March 26-28, Stanford University, Stanford

DELAIS-ROUSSARIE E., YOO H., (2014). Rythme et synthèse de la parole : études comparées des patrons rythmiques de différents genres. *Nouveaux Cahiers de Linguistique Française* 31.pp. 237-247 http://www.llf.cnrs.fr/fr/node/4927

DE LOOZE C., HIRST D., (2014). The OMe (Octave-Median) scale: A natural scale for speech melody, In 7th *Speech Prosody*, 10.21437/SpeechProsody.2014-170

DOUKHAN, D. (2007). *Synthèse de parole expressive au-delà du niveau de la phrase : le cas du conte pour enfant: conception et analyse de corpus de contes pour la synthèse de parole expressive*, Thèse de doctorat, Université Paris 11.

DOUKHAN D, RILLIARD A, ROSSET S, ADDA-DECKER M, D'ALESSANDRO C., (2011). Prosodic Analysis of a Corpus of Tales, *In Proceedings INTERSPEECH*

ETSI ES 202 212 V1.1.1, STQ (2005). Distributed speech recognition; Extended advanced front-end feature extraction.

GELIN R., D'ALESSANDRO C., LE Q., DEROO O., DOUKHAN D., MARTIN J., PELACHAUD C., RILLIARD A., and ROSSET S., (2010). Towards a storytelling humanoid robot. In *AAAI Fall Symposium Series on Dialog with Robots*, *(pp 137-138)*

GRANSTRÖM, B., & HOUSE, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication, 46(3-4), (pp. 473-484)*

INANOGLU, Z., & YOUNG, S. (2009). Data-driven emotion conversion in spoken English. *Speech Communication*, *vol. 51, no. 3, (pp. 268–283)*

SARKAR P, HAQUE A, KUMAR DUTTA A, REDDY M G., HARIKRISHNA D M, PRASENJIT D., RASHMI V., NARENDRA N P, SUNIL Kr. S B, JAINATH YADAV K., RAO S., (2014). Designing Prosody Rule-set for Converting Neutral TTS Speech to storytelling style speech for Indian Languages: Bengali, Hindi and Telugu *In* 2014 Seventh International Conference on Contemporary Computing (IC3)

SCHERER. K.R. (2005) What are emotions? And how can they be measured*? Social Science Information*, vol. 44(4), *(pp 695-729)*

TAO, J., KANG, Y., & LI, A. (2006). Prosody conversion from neutral speech to emotional speech. *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 14, no. 4, *(pp.1145-1154)*

VAUDABLE Ch., & DEVILLERS L. (2012). Negative emotions detection as an indicator of dialogs quality in call centers, *In Proceedings 37th ICASSP*, Kyoto, Japan,