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

Segmentation automatique en périodes pour le français parlé

Natalia Kalashnikova[1], Iris Eshkol-Taravella[2], Loïc Grobol[3,4], François Delafontaine[1]

(1) LLL UMR 7270, 10 Rue de Tours, 45065 Orléans, France
(2) MoDyCo UMP 7114, 200 Avenue de la République 401B, 92001 Nanterre, France
(3) Lattice, 1 Rue Maurice Arnoux, 92120 Montrouge, France
(4) LLLF (Université de Paris, CNRS)

RÉSUMÉ

Nous proposons la comparaison de deux méthodes de segmentation automatique du français parlé en périodes macro-syntaxiques, qui permettent d'analyser la syntaxe et la prosodie du discours. Nous comparons l'outil Analor (Avanzi *et al.*, 2008) qui a été développé pour la segmentation des périodes prosodiques et les modèles de segmentations utilisant des CRF et des traits prosodiques et / ou morpho-syntaxiques. Les résultats montrent qu'Analor divise le discours en plus petits segments prosodiques tandis que les modèles CRF détectent des segments plus larges que les périodes macro-syntaxiques. Cependant, les modèles CRF ont de meilleurs résultats qu'Analor en termes de F-mesure.

ABSTRACT

Automatic Period Segmentation of Oral French

Natural Language Processing in oral speech segmentation is still looking for a minimal unit for analyze. In this work, we propose a comparison of two methods of automatic segmentation in macro-syntactic periods which allows to take into account syntactic and prosodic components of speech. We compare the performances of an existing tool Analor (Avanzi *et al.*, 2008) developed for automatic segmentation of prosodic periods and of CRF models relying on syntactic and / or prosodic features. We find that Analor tends to divide speech into smaller segments and that CRF models detect larger segments than macro-syntactic periods. However, in general CRF models perform with better results than Analor in terms of F-measure.

MOTS-CLÉS : français oral, segmentation automatique, périodes, CRF, unités macro-syntaxiques.

KEYWORDS: spoken language, automatic segmentation, period, oral french, CRF, macro-syntactic units.

1 Introduction

Dans le domaine du traitement automatique des langues, la segmentation des données linguistiques est une étape préalable à la plupart des tâches. Pour le traitement du langage écrit, l'unité de base est la phrase. Or ce type de segmentation est d'une pertinence limitée pour le langage parlé. C'est la raison pour laquelle Lacheret & Victorri (2002) proposent une unité de segmentation appelée la période prosodique. Cette notion est fondée sur les observations et les analyses de l'oral. Ainsi, cette approche ne tient pas compte de la syntaxe ni de la sémantique.

Volume 2 : Traitement Automatique des Langues Naturelles, pages 241–248.
HAL : hal-02784773.

Notre travail s'inscrit dans le cadre du projet SegCor dont le but est le développement de plusieurs outils de segmentation automatique des unités linguistiques. Dans ce travail, nous nous intéressons aux périodes dans le cadre du modèle fribourgeois de la macro-syntaxe. Ce modèle définit les périodes du point de vue prosodique et syntaxique. L'objectif est de créer un outil automatique pour la segmentation de périodes macro-syntaxiques.

Notre approche aborde la tâche de segmentation comme un problème d'étiquetage des séquences. Pour cela, nous proposons d'utiliser un algorithme d'apprentissage automatique utilisant les *Conditional Random Fields* (Lafferty *et al.*, 2001) en nous appuyant sur des traits lexicaux, syntaxiques et prosodiques.

La suite de cet article est divisée en 7 parties. Les sections 2, 3 et 4 présentent un état des lieux de la recherche sur les notions de périodes, corpus et d'annotation manuelle qui sert de référence pour les méthodes automatiques. Les sections 5, 6 et 7 décrivent les expériences d'annotation automatique, leurs résultats, la conclusion et les perspectives pour de futures recherches.

2 Notion de période

Lacheret & Victorri (2002) définissent la période comme la structure prosodique qui lie plusieurs constructions syntaxiques dans un seul bloc discursif. Plusieurs périodes prosodiques peuvent aussi être incluses dans une seule structure syntaxique. Les périodes sont définis selon les paramètres prosodiques suivants : 1) Pause d'au moins 300 millisecondes ; 2) Différence de hauteur entre la valeur moyenne de la F0 de tout le signal acoustique et la dernière valeur de la F0 avant la pause ; 3)Différence de hauteur entre la dernière valeur de la F0 avant la pause et la première valeur de la F0 après la pause ; 4) Absence des signes d'hésitation (« euh ») avant et après la pause.

(1) et vous logez euh () la le la façade du théâtre (0.72)

Dans l'exemple (1) la durée de la pause marque la fin de la période après le mot "théâtre". Analor (Avanzi *et al.*, 2008) est un outil semi-automatique développé dans le cadre de cette théorie.

Une autre approche est celui de Berrendonner (2012) qui considère les périodes comme une unité prosodique autonome définie par son contour mélodique conclusif (Berrendonner, 2017). Les approches macro-syntaxiques s'appuient sur la prosodie pour analyser la structure syntaxique de l'oral (Blanche-Benveniste *et al.*, 1990; Cresti *et al.*, 2011). Pour cette raison, la période constitue potentiellement à la fois la structure complète et l'unité maximale de monologue (Blanche-Benveniste, 2012; Berrendonner, 2012, 34-35). A notre connaissance il n'existe aucun outil pour la segmentation automatique des périodes macro-syntaxiques.

Notre étude vise à trouver la méthode la plus performante pour la segmentation automatique des périodes macro-syntaxiques. Dans ce but, nous comparons deux méthodes. La première utilise Analor, qui ne nécessite pas d'entraînement et qui n'analyse pas la syntaxe des périodes. La deuxième reformule cette segmentation comme une tâche d'étiquetage de séquences — une modélisation déjà utilisée avec succès pour d'autres tâches de segmentation en français (Tellier *et al.*, 2012, 2014; Eshkol-Taravella *et al.*, 2019; Tellier *et al.*, 2013). Pour cette deuxième méthode, nous nous appuyons sur un algorithme d'apprentissage automatique bien connu : les CRF (*Conditional Random Fields*) (Lafferty *et al.*, 2001) en utilisant des traits syntaxiques et prosodiques.

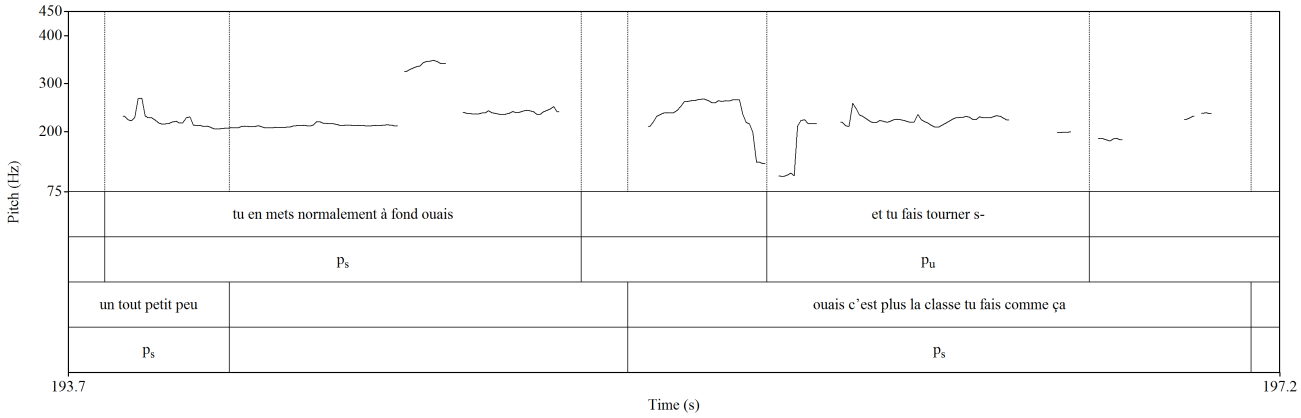

FIGURE 1 – Exemple de courbe intonative.

3 Corpus

Dans cette étude, nous travaillons sur un corpus pilote composé de 10 transcriptions de dialogues de 10 minutes et de 1 transcription de monologue de 20 minutes. Les extraits ont été sélectionnés pour représenter les différents types de discours du point de vue de l'environnement conversationnel, des relations entre les locuteurs, etc. Le corpus pilote est manuellement annoté en périodes macro-syntaxiques (voir Section 4).

4 Annotation manuelle

L'annotation manuelle réalisée par un annotateur en utilisant le logiciel Praat (Boersma & Weenik, 2002) s'est appuyée sur la définition de la période proposée par le modèle de la macro-syntaxe fribourgeoise (Berrendonner, 2012). L'un de ces deux critères devait être rempli pour reconnaître une frontière de période : la détection d'un contour intonatif conclusif ou un changement effectif de locuteur.

La détection de contours intonatifs a reposé sur une approche perceptive, par l'écoute du locuteur. L'annotation manuelle suit les mêmes propriétés dégagées par la théorie de Lacheret & Victorri (2002) qui servent à l'annotation automatique. Quant au changement de locuteur, il s'agit soit de cas où le locuteur interrompt son discours du fait de l'intervention d'un interlocuteur, soit de cas où le locuteur abandonne une structure incomplète après une longue pause (plus de 0.8 seconde).

(2) ELI tu en mets normalement à fond ouais (0.5)

BEA ouais [c'est plus classe tu fais] comme ça

ELI [et tu fais tourner]

La figure 1 illustre l'exemple (2) et le complète avec la courbe de la fréquence fondamentale.

La 1ère période d'ELI est suivie d'un changement de locuteur accompagné par la perception d'un contour conclusif montant que la mesure acoustique échoue à reproduire. La seconde période d'ELI est interrompue par le chevauchement, sans contour conclusif mais suivie d'une longue pause. La

période de BEA telle que reportée dans l'exemple 2 présente par ailleurs une structure composée de multiples unités micro-syntaxiques.

Au-delà du changement de locuteur, la syntaxe a également joué un rôle prédominant dans la gestion des pauses qui peuvent être soit une simple suspension du discours, soit l'abandon même momentané du tour par le locuteur, par exemple pour une requête lexicale (Lerner, 1991). Cette distinction devient essentielle en contexte monologique, y compris pour des pauses inférieures à 0.8 seconde. Dans de tels cas les indices prosodiques seuls s'avèrent insuffisants.

5 Expériences

5.1 Analor

Le corpus-pilote est d'abord annoté semi-automatiquement avec l'outil Analor (Avanzi *et al.*, 2008). Cet outil prend en compte les paramètres prosodiques développés dans le cadre de la théorie de Lacheret & Victorri (2002). Il semble important d'explorer la performance de l'outil pour l'annotation des périodes macro-syntaxiques due à la différence des définitions des périodes.

5.1.1 Pré-traitement

Le lancement du processus d'annotation sur Analor nécessite les fichiers PitchTier et TextGrid (format Praat). Les fichiers TextGrid contiennent 3 tires : token (un intervalle par token de l'enregistrement), locuteur (un intervalle par tour de parole) et l'annotation manuelle des périodes (un intervalle par période pour chaque locuteur).

5.1.2 Expériences

Analor est un outil pré-entraîné, on peut donc utiliser toutes les données du corpus-pilote pour la segmentation. Après la procédure d'annotation, Analor crée un autre fichier TextGrid qui contient une nouvelle tire de l'annotation automatique pour chaque fichier son. Analor réalise l'étiquetage des périodes sur une seule tire, tandis que le fichier TextGrid de l'annotation manuelle contient une tire par locuteur pour chaque fichier son. Nous avons résolu ce problème par la séparation manuelle de cette tire en une tire par locuteur. Un autre problème rencontré est le nombre différent de périodes entre l'annotation manuelle et l'annotation automatique. La solution pour cela est la tokenisation des périodes de l'annotation manuelle et de l'annotation automatique.

5.2 Modèles CRF

Les Conditional Ramdom Fields (CRF, parfois « Champs Markoviens Conditionnels » en français), (Lafferty *et al.*, 2001) sont des modèles d'étiquetage de séquences conçus pour permettre un apprentissage automatique robuste vis-à-vis d'effets à longue distance. En particulier, leur usage pour des tâches d'étiquetage modélisant une segmentation est bien connu pour des tâches de chunking

(Eshkol-Taravella *et al.*, 2019; Tellier *et al.*, 2012) et la reconnaissance d'entités nommées (Dupont & Tellier, 2014).

5.2.1 Traits

Pour le développement des modèles CRF nous utilisons deux types de traits : prosodiques et morpho-syntaxiques. Les traits prosodiques sont la fréquence fondamentale (les valeurs maximale, minimale et moyenne), l'intensité (les valeurs maximale, minimale et moyenne) et la durée pour chaque token.

Les traits morpho-syntaxiques sont les POS étiquetés par TreeTagger (Schmid, 1994) et les chunks issus d'un outil développé par Eshkol-Taravella *et al.* (2019) pour le projet SegCor. Ainsi, on construit 3 types de modèles CRF : le premier est créé uniquement sur les traits prosodiques, le deuxième est basé sur les traits prosodiques et morpho-syntaxiques et le troisième est entraîné uniquement sur les traits morpho-syntaxiques.

Nous faisons cette répartition dans le but de répondre à la question suivante : est-ce que les traits prosodiques contiennent assez d'information pour réaliser la segmentation des périodes macro-syntaxiques ou avons-nous aussi besoin des traits morpho-syntaxiques ?

5.2.2 Pré-traitement

TABLE 1 – Exemple de traits prosodiques et d'étiquettes pour le modèle CRF

mot	$f_0^{\max}$	f_0^{mean}	$f_0^{\min}$	durée	$\text{int}_{\max}$	int_{mean}	$\text{int}_{\min}$	BILU
ça	9	10	9	82	40	41	39	pA_B
va	8	9	7	83	41	43	39	pA_L
dis	9	14	6	82	41	41	40	pA_B
je	13	14	10	81	42	43	41	pA_I
voulais	8	10	7	81	42	43	42	pA_I
te	7	9	6	80	43	43	42	pA_I
demander	9	10	6	86	42	42	42	pA_L
demain	9	14	6	84	39	40	38	pA_B

Pendant la phase de pré-traitement, les traits acoustiques sont extraits pour chaque token en utilisant le logiciel Praat. Les valeurs prosodiques sont divisées en groupes de valeurs pour faciliter l'entraînement des modèles CRF. Les valeurs d'intensité sont discrétisées avec un pas de 10, de la F0 avec un pas de 20 et de la durée avec un pas de 0.1. Les fichiers TextGrid contenant les tokens et les traits morpho-syntaxiques sont convertis au format tabulaire.

Les données sont organisées selon les séquences de tours de parole. Un tour de parole peut donc contenir plusieurs périodes. En tenant compte du fait que le corpus pilote contient peu de données, notamment des tours de parole avec plusieurs périodes, nous avons décidé d'élargir le corpus en multipliant les données existantes. On a gardé les mêmes valeurs de traits mais en remplaçant les tokens par des faux mots. Les faux mots ont été créés par tirage aléatoire des caractères. Cela a permis au système d'avoir plus de données prosodiques pour l'entraînement tout en l'empêchant de procéder par simple mémorisation des mots.

Nous analysons également l'influence de l'intensité et de la fréquence fondamentale en entraînant les modèles uniquement sur un de ces traits prosodiques. Cette méthode permet d'observer quel paramètre prosodique est le plus important pour la segmentation des périodes.

Le corpus est divisé en 3 parties : 60 % pour l'entraînement, 30 % pour l'évaluation et 10 % pour le développement. Au total, nous avons 6 configurations différentes pour entraîner nos modèles CRF.

6 Expériences et résultats

Nous utilisons le logiciel Wapiti (Lavergne *et al.*, 2010) pour construire les modèles CRF.

TABLE 2 – Comparaison des résultats avec les différents traits.

Modèle	P	R	F
Analor	0.52	0.22	0.31
Prosodie	0.56	0.78	0.66
Morphosyntaxe	0.54	0.68	0.60
Prosodie + morphosyntaxe	0.56	0.78	0.66
Prosodie + morphosyntaxe + augmentation	0.56	0.70	0.62
f_0	0.55	0.76	0.64
Intensité	0.72	0.55	0.62

Le tableau 2, compare les performances de nos deux méthodes en termes de précision, de rappel et de F-mesure, calculés à partir des étiquettes BILU des périodes annotées automatiquement et manuellement.

Les valeurs rapportées sont des valeurs de détection des périodes (et non simplement des étiquettes) en considérant qu'une période est détectée si le modèle a correctement identifié ses frontières gauches et droites.

Pour Analor, le score de précision est plus haut que le score de rappel. Ceci est dû au fait qu'Analor détecte des segments plus petits que les périodes macro-syntaxiques. De plus, dans la plupart des cas, les meilleurs résultats correspondent aux locuteurs ayant le moins de temps de conversation et inversement.

Pour les modèles CRF, il semble que bien que l'utilisation uniquement des traits morpho-syntaxiques donne déjà de meilleurs résultats que l'utilisation directe d'Analor, les ajouter à un modèle ayant accès aux traits prosodiques n'améliore pas les performances, les traits prosodiques seuls obtenant déjà les meilleurs résultats.

Pour définir l'importance de chaque trait prosodique, nous comparons les résultats des modèles construits uniquement sur les valeurs de la F0 et de l'intensité. Les résultats obtenus montrent une grande complémentarité entre ces traits, chacun contribuant à un des aspects de la détection, et leur combinaison donnant de meilleurs résultats que leurs usages en isolation.

7 Conclusions et perspectives

Dans ce travail, nous avons présenté une nouvelle méthode pour la segmentation automatique de l'oral. Nous avons analysé les périodes macro-syntaxiques qui permettent de tenir compte du contenu prosodique et morpho-syntaxique du discours. La performance d'Analor n'est pas assez satisfaisante pour l'annotation des périodes macro-syntaxiques.

Tous les modèles CRF ont montré de meilleurs résultats qu'Analor. La F-mesure varie entre 0,54 et 0,66 parmi les modèles CRF différents. Si l'on compare la performance de chaque modèle CRF et que l'on tient compte du temps de pré-traitement, le modèle le plus performant est développé sur le corpus initial de la 1ère échelle des valeurs.

Nous avons également trouvé que le modèle fondé sur les traits de la F0 montre de meilleurs résultats que le modèle développé sur les traits de l'intensité.

Étant donné la quantité limitée des données, il serait intéressant de procéder à une validation croisée lors de l'entraînement des modèles. De plus, on pourrait appliquer les tests de significativité sur les résultats obtenus par les modèles.

Une piste de recherche pourrait être une évaluation de l'importance des différentes caractéristiques prosodiques sur la performance des modèles CRF. Il serait également envisageable de calculer la différence entre les valeurs d'un mot et du mot précédent lors du pré-traitement.

Références

AVANZI M. (2005). Quelques hypothèses à propos de la structuration interne des périodes. In *Actes Du Symposium Interface Discours-Prosodie*, Aix-en-Provence, France.

AVANZI M. (2012). *L'interface prosodie/syntaxe en français : dislocations, incises et asyndètes*. GRAMM-R. Bruxelles, Belgique : Peter Lang.

AVANZI M., LACHERET A. & VICTORRI B. (2008). Analor, un outil d'aide pour la modélisation de l'interface prosodie-grammaire. *Travaux linguistiques du CerLiCo*, **21**, 27–46.

BALTHASAR L. & BERT M. (2005). La plate-forme "Corpus de langues parlées en interaction" (CLAPI) : historique, états des lieux, perspectives. *LIDIL - Revue de linguistique et de didactique des langues*, **31**, 13–33.

BAUDE O. & DUGUA C. (2011). (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? *Corpus*, **10**, 99–118. HAL : hal-01162479.

BERRENDONNER A., Éd. (2012). *Grammaire de La Période*. Sciences pour la communication. Peter Lang. DOI : 10.3726/b11424.

BERRENDONNER A. (2017). La notion de période (note terminologique). *Encyclopédie grammaticale du français*.

BLANCHE-BENVENISTE C. (2012). Postface. In (Berrendonner, 2012), p. 341–355. DOI : 10.3726/b11424.

BLANCHE-BENVENISTE C., BILGER M., ROUGET C., VAN DEN EYNDE K. & WILLEMS D. (1990). *Le Français parlé : études grammaticales*. Paris, France : Centre national de la recherche scientifique.

BOERSMA P. & WEENIK D. (2002). Praat, a system for doing phonetics by computer. *Glot International*, **5**(9/10), 341–345.

CRESTI E., MONEGLIA M. & TUCCI I. (2011). Annotation de l'entretien d'Anita Musso selon la Théorie de la langue en acte. *Langue francaise*, **170**(2), 95–110.

DUPONT Y. & TELLIER I. (2014). Un reconnaisseur d'entités nommées du Français. In *Actes de la 21e conférence sur le Traitement Automatique des Langues Naturelles*, volume 3, p. 40–41, Marseille, France : Association pour le Traitement Automatique des Langues.

ESHKOL-TARAVELLA I., BAUDE O., MAUREL D., HRIBA L., DUGUA C. & TELLIER I. (2011). Un grand corpus oral « disponible » : Le corpus d'Orléans 1 1968-2012. *Traitement Automatique des Langues*, **53**(2), 17–46.

ESHKOL-TARAVELLA I., MAAROUF M., BADIN F. & SKROVEC M. (2019). Chunker différents types de discours oraux : défis pour l'apprentissage automatique. In *Actes de La 26eConférence sur le Traitement Automatique des Langues Naturelles*, Toulouse, France : ATALA.

LACHERET A. & VICTORRI B. (2002). La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques. *Verbum*, **24**, 55–72.

LAFFERTY J., MCCALLUM A. & PEREIRA F. (2001). Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data. In *18th International Conference on Machine Learning*, ICML '01, p. 282–289, San Francisco, CA, USA : Morgan Kaufmann.

LAVERGNE T., CAPPÉ O. & YVON F. (2010). Practical Very Large Scale CRFs. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, p. 504–513, Uppsala, Sverige : Association for Computational Linguistics.

LERNER G. H. (1991). On the syntax of sentences-in-progress*. *Language in Society*, **20**(3), 441–458. DOI : 10.1017/S0047404500016572.

SCHMID H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*, Manchester, UK.

TELLIER I., DUCHIER D., ESHKOL I., COURMET A. & MARTINET M. (2012). Apprentissage automatique d'un chunker pour le français. In G. S. GEORGES ANTONIADIS, HERVÉ BLANCHON, Éd., *Conférence Conjointe JEP-TALN-RECITAL 2012*, volume 2, p. 431–438, Grenoble, France.

TELLIER I., DUPONT Y., ESHKOL I. & WANG I. (2013). Adapt a Text-Oriented Chunker for Oral Data : How Much Manual Effort Is Necessary ? In H. YIN, K. TANG, Y. GAO, F. KLAWONN, M. LEE, T. WEISE, B. LI & X. YAO, Éds., *Proceedings of the 14th International Conference on Intelligent Data Engineering and Automated Learning*, Lecture Notes in Computer Science, p. 226–233, Berlin, Heidelberg : Springer. DOI : 10.1007/978-3-642-41278-3_28.

TELLIER I., ESHKOL-TARAVELLA I., DUPONT Y. & WANG I. (2014). Peut-on bien chunker avec de mauvaises étiquettes POS ? In B. BIGI, Éd., *Actes de la 21e conférence sur le Traitement Automatique des Langues Naturelles*, p. 125–136, Marseille, France.