
Description de la morphosyntaxe du verbe aller dans le roman Pélégie-la-charrette d'Antonine Maillet, à l'aide du logiciel INTEX

Aline Chapados

Laboratoire d'Analyse de Données Textuelles, (LADT), Université de Moncton
[Aline_c@email.com](mailto:aline_c@email.com)

Gisèle Chevalier

Laboratoire d'Analyse de Données Textuelles, (LADT), Université de Moncton
chevalg@umoncton.ca

Sylvia Kasparian

Laboratoire d'Analyse de Données Textuelles, (LADT), Université de Moncton
kaspars@umoncton.ca

ABSTRACT. The use of the verb “aller” (‘to go’) in regional french often differs from standard French, including different contexts of use and particular forms of the verb. Our project has two goals: first, to find the different morphosyntactic contexts of the verb “aller” in Acadian French spoken in New Brunswick, using INTEX programs; second to integrate the results into software dictionaries following the methodological framework developed by Maurice Gross of the LADL. This work represents the first stage in a larger project to develop an electronic dictionary of Acadian French with INTEX. In this paper we will present the different graphic forms of the verb that were found in the oral and written corpora and also the graphs built for the description of the morphosyntax of the word. Finally we will present the dictionary of frozen expressions with the verb “aller” found in our corpora.

KEYWORDS : INTEX, Acadian French, the verb ‘to go’, frozen expressions, lexeme-grammar based syntax.

RÉSUMÉ. Les utilisations du verbe aller dans les français régionaux diffèrent de celles du français standard. Nous y retrouvons des contextes d'utilisation ainsi que des formes de verbe particulières. Ce projet vise deux objectifs : d'abord, dépouiller, par l'intermédiaire de l'outil informatique INTEX, les différents contextes morphosyntaxiques du verbe aller, dans le français acadien du Nouveau-Brunswick, et ensuite, intégrer ces résultats au logiciel sous la forme d'un dictionnaire, dans le cadre méthodologique développé au LADL par Maurice Gross. Ce travail se veut un premier pas vers la création d'un dictionnaire Intex du français acadien. Nous présenterons les formes graphiques du verbe aller que l'on retrouve dans les corpus oraux et écrits ainsi que les graphes construits dans Intex pour les décrire. Nous aborderons enfin le dictionnaire des expressions figées du verbe aller construit à partir de ces données

MOTS-CLÉS : INTEX ; français acadien ; verbe aller ; expressions figées, lexique-grammaire

1. Introduction

L'informatique a ouvert des voies jusqu'à aujourd'hui inimaginables en analyse de corpus avec l'arrivée des concordanciers qui exécutent automatiquement un travail de défrichage des

textes. Ce travail était auparavant l'œuvre d'une carrière, car il devait être fait manuellement. Du simple calcul des mots par la reconnaissance graphique des formes, on est passé à des résultats un peu plus fins grâce au processus de lemmatisation qui met en relation les différentes formes graphiques d'un même mot comme les variantes du singulier et du pluriel (*croyance/ croyances; gouvernemental/ gouvernementaux*), les variantes orthographiques (*événement/ évènement*) et les multiples variantes d'un verbe conjugué aux différents temps, modes et personnes (*vais, irai, irais, allais, va, allons, allez, vont, aillent...*). Ce premier travail ne résout pas la question de l'ambiguïté des formes (*la* : déterminant, pronom, nom d'une note de musique, et dans certains logiciels, adverbe : *là*) et encore moins de l'ambiguïté sémantique (femme vs homme; mari; fille). Mais au stade de la pré-analyse, on peut dire que plus les textes du corpus à traiter se conforment aux normes de la langue écrite standard, meilleurs seront les taux de reconnaissance des formes et leur catégorisation grammaticale. Les formes non-reconnues pourront être portées au compte des idiosyncrasies qui relèvent d'un vocabulaire technique, d'usages régionaux de la langue ou de phénomènes d'oralité, comme on en rencontre couramment dans le discours rapporté, tant dans les textes littéraires que journalistiques.

Lorsqu'on travaille sur un corpus de textes marginaux de ce point de vue, les soi-disant « scories » se présentent de façon récurrente et deviennent la matière première à organiser. Il est alors souhaitable d'enregistrer les observations dans des lexiques ou des grammaires particulières de façon à les appliquer automatiquement à d'autres textes du même corpus. C'est le projet qui occupe une équipe de chercheurs décrivant le corpus de productions orales et écrites produits en Acadie dans une perspective littéraire ou linguistique¹. Le présent article présente la première étape de la construction d'une première parcelle d'un dictionnaire électronique des parlers acadiens à l'aide de l'outil INTEX développé par Max Silberztein (1993). Le travail se concentre sur un phénomène ponctuel : la description morpho-syntaxique des manifestations du verbe *aller*, dans le roman intitulé *Pélagie-la-Charette* (1979) de l'écrivain acadien Antonine Maillet.

2. Description de l'outil

INTEX est un analyseur linguistique qui fonctionne au niveau de la phrase, donc de la combinatoire entre les mots et groupes de mots qui constituent la phrase. Il est fondé sur les principes de la grammaire transformationnelle harrisienne appliqués au modèle de description linguistique développé par Maurice Gross (1975, 1982 ; [BOONS 76]) et pratiqué sans relâche au LADL (Laboratoire d'automatique documentaire et linguistique). Selon ce modèle, l'unité de base de la signification est la phrase élémentaire, c'est-à-dire le verbe avec son sujet et ses compléments (le prédicat et les arguments). L'entrée du dictionnaire doit donc énumérer les propriétés syntaxiques, fonctionnelles et sémantiques du mot. Le lexique-grammaire est une composante particulière du modèle dans laquelle sont décrits les traits morphologiques, syntaxiques et sémantique qui régissent la combinatoire entre les prédicats de la langue et les arguments qu'ils sélectionnent pour former des phrases grammaticales et sémantiquement bien formées. Le modèle de M. Gross a été formalisé dès les années 60 et mis à la disposition des chercheurs grâce au logiciel d'analyse linguistique INTEX.

INTEX permet à la fois de décrire une langue par la localisation de constructions (patterns) au moyen d'expressions ou de graphes, et de faire la reconnaissance automatique de textes dans les langues pour lesquelles ont été élaborés des dictionnaires (dictionnaire de mots

¹ L'Acadie est un espace culturel et linguistique implanté à l'époque coloniale française et se situe géographiquement aujourd'hui dans les Provinces Atlantiques du Canada, territoire qu'elle partage avec la communauté anglophone, dominante aux plans démographique et politique.

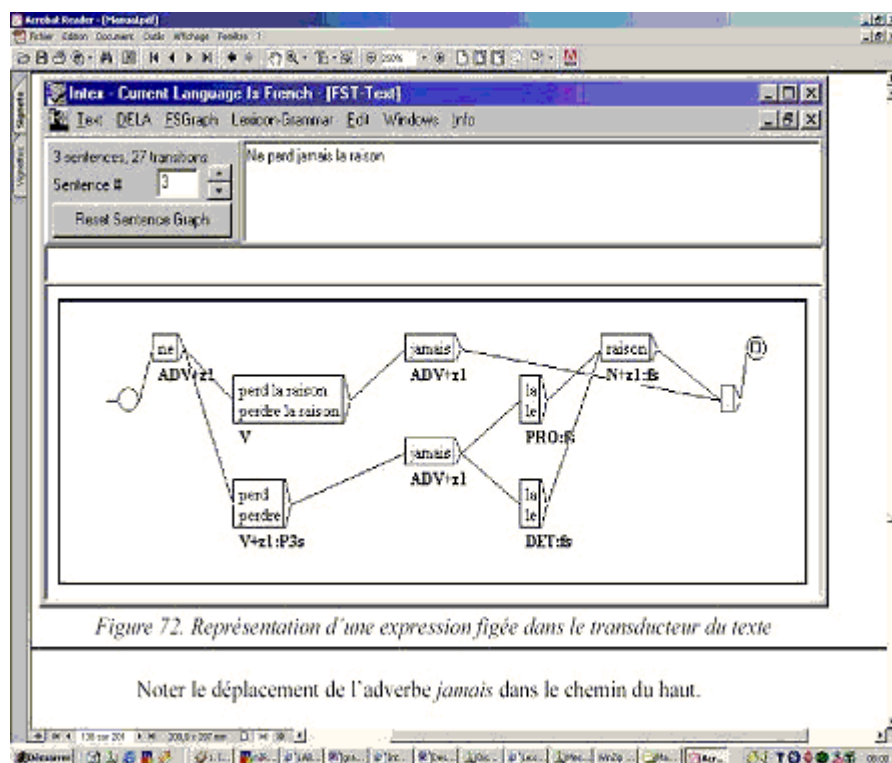


Figure 2. Graphe de l'expression figée « perdre la raison ».

Notre contribution consiste à décrire le travail d'analyse préalable à la constitution des tables et à la formalisation. Nous nous en tenons à une utilisation particulière de l'outil INTEX, l'identification des expressions présentes dans le corpus et leur description, sans chercher encore à formaliser les résultats.

3. Le corpus

Le corpus utilisé est un ouvrage littéraire d'un auteur acadien originaire de Bouctouche, dans le sud-est du Nouveau-Brunswick. C'est la narration de l'Odyssée des familles acadiennes déportées en Nouvelle-Angleterre, mues par l'espoir de regagner leur pays et les terres qui leur ont été confisquées, bref, de retrouver la vie idyllique d'avant le « Grand Dérangement » (1755-1762). Le texte est rédigé dans un français standard soutenu –il a valu le prix Goncourt à Madame Mailliet en 1979, –et parsemé d'expressions en « vieil acadien », mots de vocabulaire, phraséologie, expressions idiomatiques qui reflètent un mode de vie périmé et des usages attribués « aux anciens », le prototype étant la finale en –ons/ont du verbe : *j'avions, ils disont*. Bref, des faits d'oralité qui se trouvent dans les dialogues ou qui sont intégrés à la narration, notamment dans le discours rapporté indirect.

4. Les transducteurs finis

Une particularité des textes qui reproduisent le parler oral acadien est la présence de formes qui ne sont pas conformes au français écrit standard. Lorsqu'on demande à INTEX de retrouver toutes les formes du verbe *aller* dans un texte acadien, il ne reconnaîtra pas par exemple la forme du subjonctif présent non standard transcrit de la façon suivante : *I faut que j'ialle*, pas plus que la deuxième personne du singulier *tu ialles* ou la troisième personne du pluriel *i(ls) iallent*. Il y a dans INTEX des éléments nommés des « transducteurs finis » qui ont été utilisés, entre autres choses, pour retrouver, des mots qui ont des graphies différentes, comme 'tsar', 'czar', au singulier ou au pluriel. Nous avons construit sur ce modèle un

transducteur qui permet à INTEX de reconnaître à la fois les variantes standard et non standard du verbe au subjonctif rencontrées en français acadien, la variante consistant en fait en l'inversion de l'ordre de « a » et « i » entre le français écrit et l'acadien parlé.

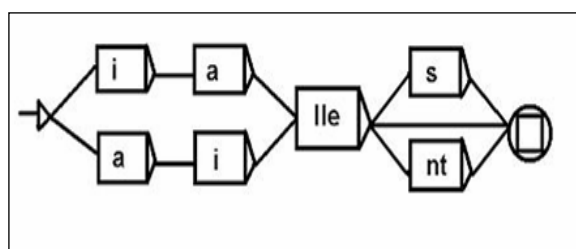


Figure 3. Transducteur des différentes graphies, en acadien, du subjonctif présent du verbe aller

La plupart des formes du verbe *aller* en français standard pouvant être utilisées dans les textes acadiens, le meilleur traitement consiste donc à appliquer le graphe de variantes acadiennes orales du subjonctif (voir la figure 4) en conjonction avec le graphe de la conjugaison du lemme <aller>, présent dans la grammaire d'INTEX.

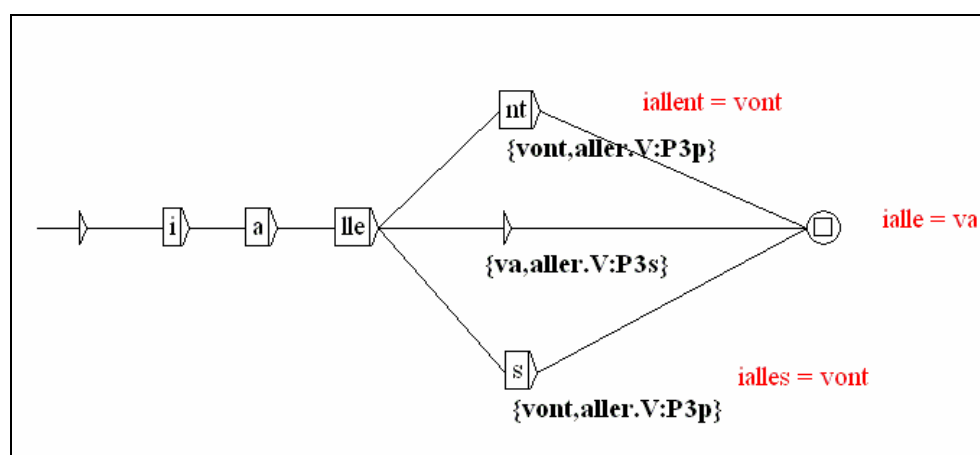


Figure 4. Transducteur des différentes formes graphiques du subjonctif de aller en acadien

5. Méthodologie

Le transducteur, appliqué avec <aller>, a repéré exactement deux cents (200) occurrences du verbe *aller* dans les 75 835 mots du roman. INTEX construit automatiquement la liste des occurrences dans les phrases où elles apparaissent. L'étape suivante consiste à produire un graphe qui puisse rendre compte de l'ensemble des constructions possibles compte tenu de la nature des éléments qui précèdent et suivent immédiatement le verbe. Les éléments ont été regroupés en fonction des catégories par approches successives des données jusqu'à épuisement des phrases attestées, ce que représente le graphe global de la figure 5.

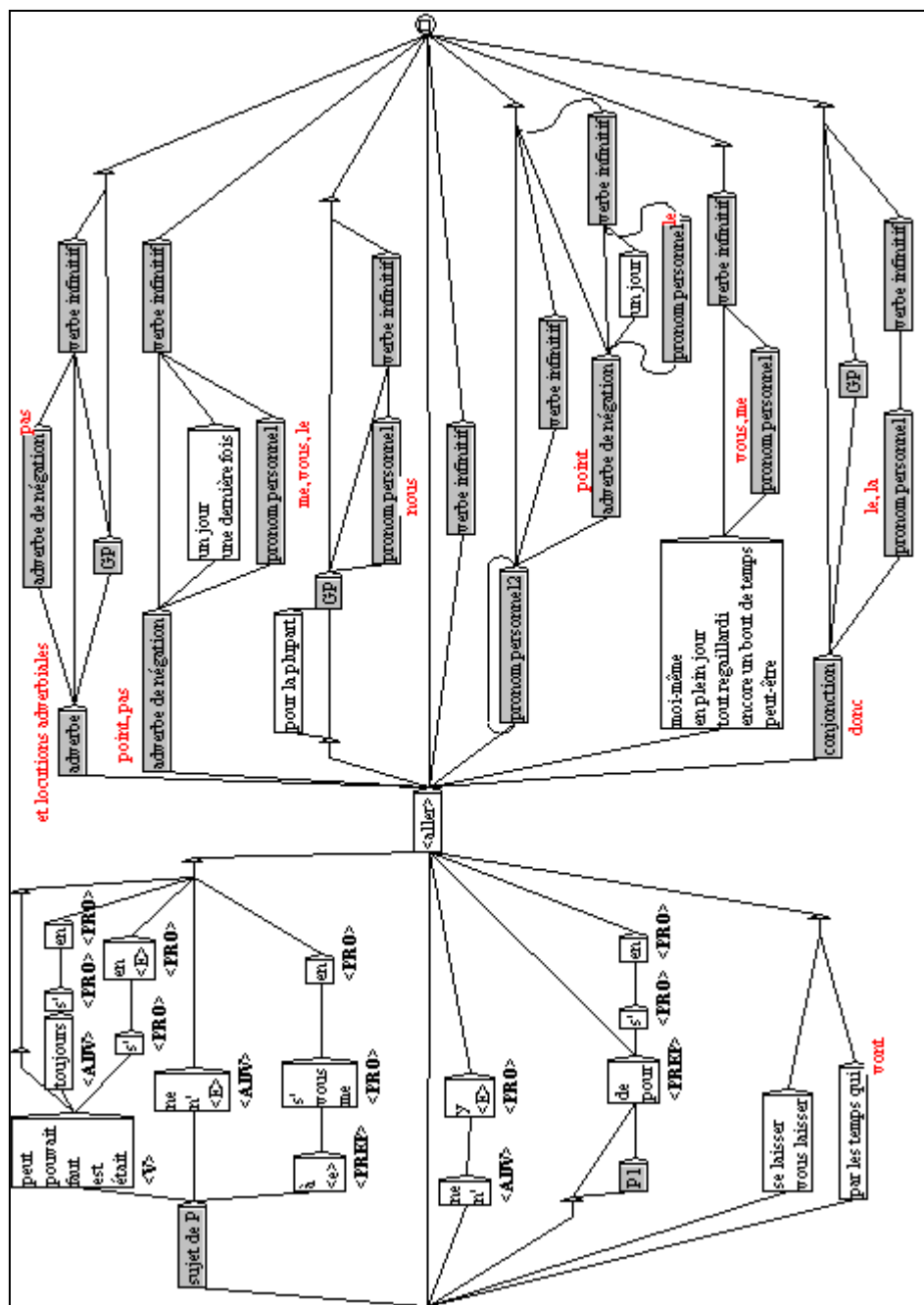


Figure 5. Transducteur du graphe global.

Ce graphe se lit de façon linéaire, de gauche à droite. Chaque chemin depuis la flèche de départ jusqu'au nœud terminal, indique un contexte d'utilisation morphosyntaxique différent du verbe *aller*, par l'auteur du livre Pélagie-la-Charrette. Passons par quatre chemins... à titre d'illustration. Le chemin minimal serait un énoncé comme *Allons!* (l'impératif ou le marqueur discursif) qui va directement de la flèche initiale au nœud terminal en ne passant que par *aller*. *Le radoteux va conter* part de [Sujet de P], traverse <aller>, poursuit son chemin par [Verbe infinitif] jusqu'au nœud terminal. *Aller* est alors semi-auxiliaire. Le chemin de *elle va sus ses vingt ans* part de [Sujet de P] (*elle*), passe par <aller> et continue à [GP] *sus ses vingt ans*, et rejoint le nœud terminal. Le dernier chemin que nous suivrons explique l'utilisation des boucles dans les graphes. Ces dernières indiquent que l'élément peut se répéter un nombre indéfini de fois. Dans *Allons-nous-en*, on part de <aller> et on passe par [pronoms personnels] pour lire *nous*, puis de nouveau pour lire *en*. On aboutit finalement au nœud terminal.

6. Description des sous-graphes

Chaque boîte ombragée indique qu'un sous-graphe a été bâti et nommé sous le nom présent dans cette case de façon à ce que INTEX le retrouve au moment de l'appliquer au texte. Le procédé est récursif, c'est-à-dire que les graphes peuvent s'emboîter les uns dans les autres. Le nom sous lequel a été sauvegardé un sous-graphe correspond toujours exactement au nom qui apparaît dans le graphe de façon à ce qu'INTEX le retrace dans la banque de graphes. Voyons par exemple, la description du graphe Sujet de P, le plus complexe de tous ceux qui sont présents dans ce travail puisqu'il comporte jusqu'à quatre niveaux d'enchâssement.

Sujet de P se compose de trois sous-graphes comme le montrent les boîtes ombragées de la figure 6 : le graphe des noms propres et des groupes nominaux qui représentent des personnes (étiquetés « +humains » [+Hum]); les groupes nominaux qui désignent de noms communs non humains (GN[-Hum]) et les pronoms sujets.

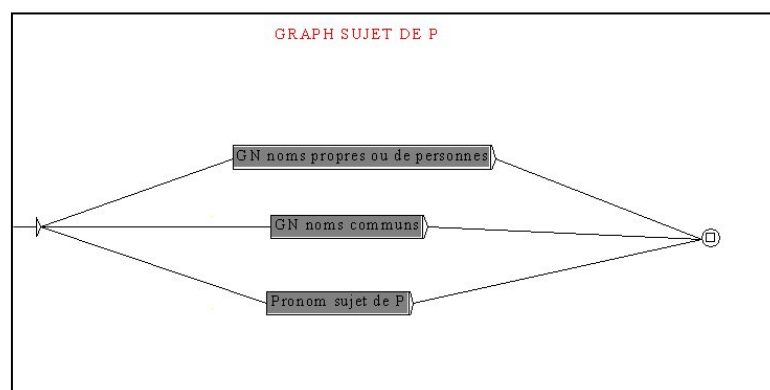


Figure 6. Transducteur du graphe « sujet de P ».

Les deux premiers sous-graphes s'ouvrent sur l'énumération des occurrences de noms propres et groupes nominaux attestés dans le corpus (voir les figures 7 et 8, à titre d'exemples).

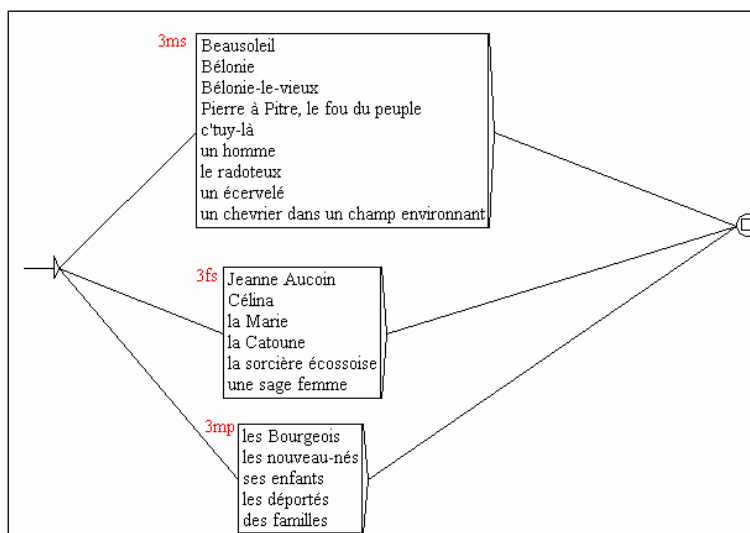


Figure 7. Transducteur du sous-graphe « GN noms propres ou GN communs de personnes [+Hum] ».

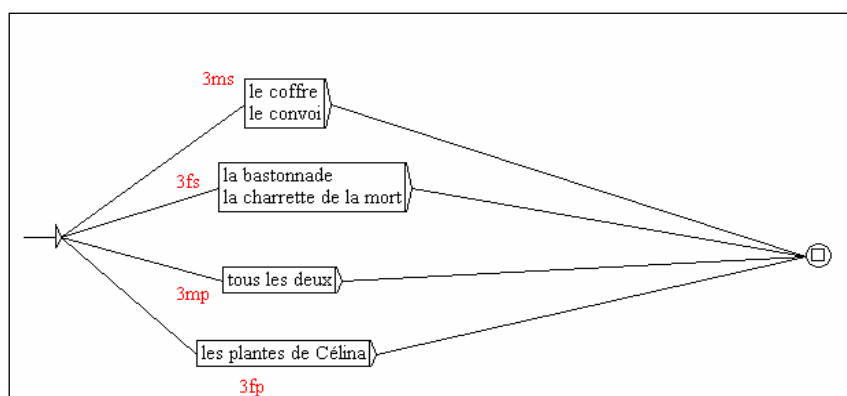


Figure 8. Transducteur du sous-graphe « GN noms communs [-Hum] »

Le troisième sous-graphe de Sujet de P intitulé « Pronoms sujets de P » dans la figure 6, comporte un nouvel ensemble de sous-graphes dont le contenu est finalement détaillé dans le graphe de la figure 9.

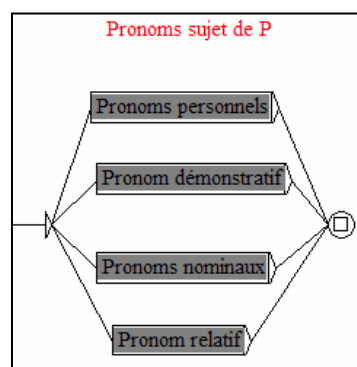


Figure 9. Transducteur du sous-graphe « Pronoms sujet de P ».

Chacune des cases de ce dernier graphe s'ouvrent sur l'énumération des pronoms sujets présents dans *Pélagie-la-Charrette*².

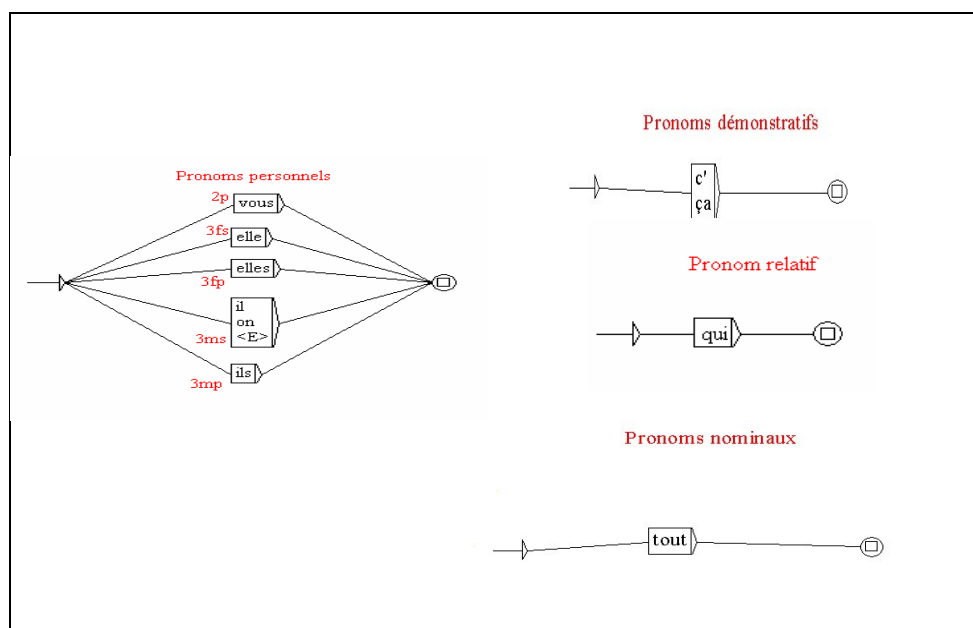


Figure 10. Transducteur des sous-graphes inclus dans « Pronoms sujet de P ».

De la même façon, les cases ombragées à droite du verbe *aller* contiennent les occurrences de verbes infinitifs, des adverbes et de groupes prépositionnels (GP) qui viennent saturer la valence du verbe *aller* dans notre corpus. Voici le graphe GP à titre d'exemple (figure 11).

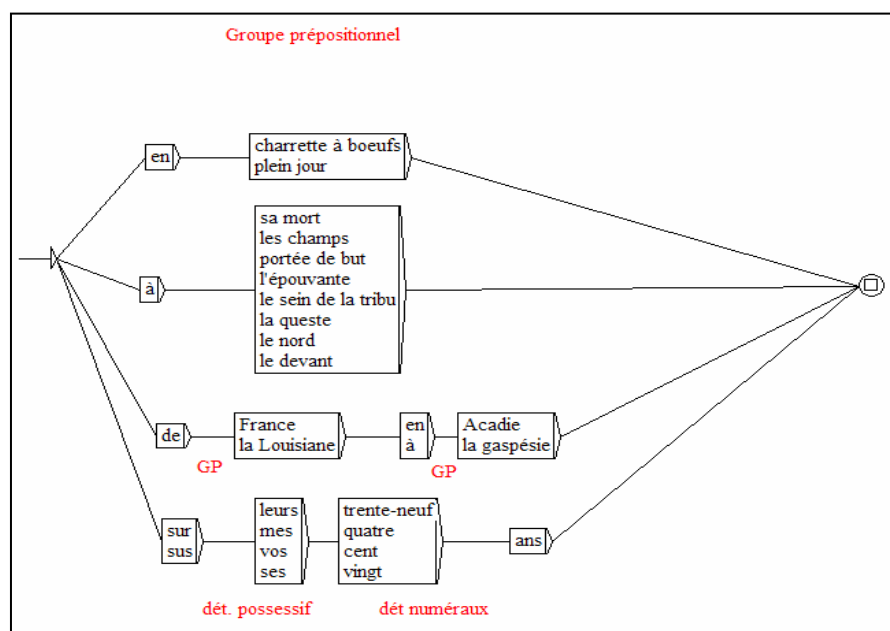


Figure 11. Transducteur du sous-graphes « GP ».

² On notera que le pronom à la deuxième personne (tu, t') et les formes orales non-standard de la troisième personne (yelle, zeux, nous-aut', vous-aut') n'apparaissent pas. Ce qui montre l'importance d'étendre la construction du Lexique-Grammaire au traitement du corpus de productions orales.

7. Observations générales

Comme nous venons de le voir, chaque case du graphe peut être composée de plusieurs autres sous-graphes, et ces sous-graphes peuvent en contenir d'autres et ainsi de suite. Le graphe de la figure 4 a le mérite de rassembler en une seule figure l'ensemble des contextes représentés dans *Pélagie-La-Charrette*, y compris les expressions figées qui sont analysées en leurs différents éléments.

Au niveau syntaxique, l'ensemble des constructions semble conforme au français standard. Le travail de catégorisation fait ressortir quelques particularismes que l'on pourra consigner dans les dictionnaires « acadiens », tels les verbes *quérir*, *corver*, la prononciation *ouère* de « voir », les adverbes *betôt* (bientôt), *itou* (aussi), le démonstratif *c'tuy-là* (celui-là), des noms propres inusités comme *Bélonie*, l'usage répandu du déterminant devant un nom propre (*la Catoune*, *la Marie*).

Mis à part les éléments contingents comme la négation et les adverbes de temps dans le contexte à droite, *aller* est semi-auxiliaire lorsqu'il est suivi d'un verbe infinitif [aller + V :inf] (1) conformément à l'usage standard. En construction libre, il est suivi d'un complément indirect [aller+GP] (2) ou d'un adverbe ou une locution adverbiale [aller+ADV] (3). Dans l'exemple (4), il est analysé comme un verbe di-transitif, soit suivi de deux compléments prépositionnels [aller+GP_{de}+GP_à]. Enfin, il est utilisé intransitivement à l'impératif (5). On peut alors se demander s'il s'agit toujours du verbe assumant son plein sémantisme, ou s'il n'est pas plutôt une interjection.

- (1) Venez, venez tous, le radoteux va conter.
- (2) J'irons au nord, mais point avant d'avoir pris des nouvelles des nôtres à Baltimore.
- (3) Bélonie est allée au-devant, en homme courtois et bien élevé, en homme fier surtout, et qui aurait le dernier mot.
- (4) La preuve, c'est que jamais en deux siècles on n'a retrouvé dans toute la ligne acadienne d'Amérique, qui va de la Louisiane à la Gaspésie, la moindre petite croix de bois [...]
- (5) ... allez! hue! et hue! et hue! Ça passe point.

On aura peut-être noté que les GP attestés dans *Pélagie-La-Charrette* (figure 11) sont pour la plupart rattachés à des expressions figées, surtout celles qui sont introduites par les préposition *à* et *sur*, comme l'attestent les exemples en (6 a-b) ci-dessous.

- (6) a) Se laisser aller à sa mort, ce n'est pas se tuer, pas tout à fait.
Voyons, vaurienne, as-tu oublié que je vas sus mes cent ans?
- b) On manquait de tout, et tout s'en allait à l'épouvante : les pailles rendaient leur paille à grands trous; les bœufs semaient entre les cailloux leurs fers usés et cobis; les gencives rongées par le scorbut saignaient et Céline, secondée par son Fou fidèle, n'arrivait pas à stopper les hémorragies.

En fait, les expressions figées suivent souvent le modèle des constructions de phrases libres, mais l'emploi de l'intransitif dans *par les temps qui vont* (7), n'est pas prévu en construction libre. L'expression n'est toutefois pas exclusive à l'acadien.

- (7) Et par les temps qui vont, une personne ne sait jamais si son adieu n'est point pour tout de bon.

Toujours comme en français standard, le parler acadien fait usage du verbe *aller* dans la construction figée pronominale composée du pronom réflexif *se* et du pronom complément de provenance *en* : [s'en aller] suivi d'une complétive infinitive (8). Une construction pronominale intéressante en acadien ressort pourtant du graphe global, celle de [s' + aller] que nous retrouvons notamment dans la phrase *Un homme peut s'aller queri'*. À mettre également dans les expressions figées particulières, la construction pronominale *s'en aller*, suivie de l'infinitif :

- (8) Vraiment, il fallait être obtus pour imaginer que les Anglais disposaient de jougs de cornes; plus obtus encore pour croire que s'ils en avaient eu, ils les auraient cédés à des parlant français; et plus fou que obtus pour s'en aller en plein jour éveiller l'ours qui dort

Le système INTEX prévoit que l'on range dans des dictionnaires différents dans les expressions figées et les utilisations libres, et qu'on consigne sous forme de graphes les différentes expressions figées en fonction de leur construction et de leur sens. Le tableau d'ensemble ne changerait pas significativement d'aspect si on extrayait les emplois figés. Seule la boîte [pronoms non sujet] dans le contexte avant <aller> est liée directement à la locution « s'en aller ». Les autres constructions ont une relation ambiguë entre la construction libre et les expressions figées.

8. Conclusion

En conclusion, ce projet a permis de bâtir un outil électronique de recherche morphosyntaxique des utilisations du verbe *aller* dans le roman *Pélagie-la-Charrette* d'Antonine Maillet (1979). Nous avons franchi un premier pas vers la description de ce verbe dans un dictionnaire électronique acadien. En se servant du graphe global obtenu, il est possible de rechercher, à l'aide d'INTEX, les structures non reconnues par ce graphe dans d'autres textes acadiens afin de les incorporer à celui-ci. On pourra alors généraliser la description et passer à la deuxième utilisation d'INTEX qui consiste à utiliser les dictionnaires et les graphes pour formaliser la langue, notamment décrire systématiquement les emplois possibles du verbe *aller* en construisant des grammaires des emplois libres et figés.

Références

[BOONS 76] BOONS, Jean-Paul, Alain GUILLET, et Christian LECLERE. *La structure des phrases simples en français. Constructions intransitives*. Présentation de Maurice Gross. Genève : Librairie Droz. p.377.

[GROSS G 94] GROSS, Gaston. « Classes d'objets et description des verbes ». *Langages*, no.115, 1994, p.15-30

[GROSS G 90] GROSS, Gaston. « Définition des noms composés dans un lexique-grammaire ». *Langue- Française*, N°87, 1990, p.84-90.

[GROSS G 84] GROSS, Gaston. « Compléments adverbiaux et verbes supports ». *Revue Québécoise de Linguistique*, N°13, 1983, p.123-156.

[GROSS M 82] GROSS, Maurice. « Une classification des phrases « figées » du français. ». *Revue québécoise de linguistique* Vol.11, N° 2, p.150-189.

[GROSS M 75] GROSS, Maurice. *Méthodes en syntaxe. Régime des constructions complétives*. Paris : Hermann, 1975. p.272.

[MAILLET 79] MAILLET, Antonine. Pélagie-la-Charrette. Montréal : Leméac, 1979. 351 pages.

[PERRONET 90] PERRONET, Louise. « Système des conjugaisons verbales dans le parler acadien du sud-est du Nouveau-Brunswick ». *Revue de l'Association de linguistique des Provinces Atlantiques*. N°12, 1990, p.81-115.

[POIRIER 97] POIRIER, Pascal. Le glossaire acadien/ Pascal Poirier ; édition critique établie par Pierre M. Gérin. Moncton : Éditions d'Acadie, 1997, 466 pages.

[SILBERZTEIN 96] SILBERZTEIN, Max. « Analyse automatique de corpus avec INTEX ». *LINX*, N°34,2, 1996, p.269-276.

[SILBERZTEIN 90] SILBERZTEIN, Max. « Dictionnaires électroniques du français ». *Langue-Française*, N°87, 1990, p.71-83.

Lien internet

<http://users.bestweb.net/~INTEX/downloads/Manuel.pdf> - page du Manuel d'utilisation d'INTEX de Max Silberztein