



**Une expérience de collaboration entre linguiste et spécialiste de TAL :  
L'exploitation du corpus CFPP 2000 en vue d'un travail  
sur l'alternance Futur simple / Futur périphrastique**

Serge Fleury & Sonia Branca-Rosoff

Sorbonne Nouvelle-Paris III

**Abstract**

This paper deals with the description on the uses of two tenses, the *futur simple* (FS) and *futur périphrastique* (FP) in a corpus of interviews with Parisians (CFPP2000). The aim is to develop the resources provided by the corpus, in other words (i) to show that interesting findings on the use of both forms can be made from observable linguistic activities and attested linguistic data and not only from introspective acceptability judgments and linguist-made examples; (ii) to show how statistical and computer software support descriptors while preventing against an uncontrolled exploitation of the obtained results. We are helped by an automatic tagger, TreeTagger, and a textometric software, *le Trameur*. Data are abundant enough (300,000 words at the time this study took place) to allow the specific value of each future to be envisaged, cases where both forms are in competition to be examined and to see if oral corpora offer support to conflicting theories on the study of these tenses. Another strength of the corpus is that it is fairly homogeneous as far as the genre is concerned (interviews are mostly conversations between 3 people, favouring relaxed exchanges). However, levels of education, jobs, ages etc. vary. The corpus therefore allows a first approach of sociolinguistic variation among native Parisians.

Cet article porte sur la description des usages de deux tiroirs verbaux, le futur simple (FS) et le futur périphrastique (FP) dans un corpus d'interviews réalisées auprès de Parisiens (CFPP2000). Il s'agit pour nous de valoriser les ressources fournies par le corpus c'est-à-dire (i) de montrer que l'on peut apprendre des choses intéressantes sur le fonctionnement des deux formes FS et FP en partant d'activités langagières observables et de faits linguistiques attestés – et pas seulement de jugements d'acceptabilité introspectifs et d'exemples "cruciaux" forgés par des linguistes. (ii) de montrer l'aide apportée par les applications informatiques et statistiques qui viennent aider les descripteurs, tout en mettant en garde contre toute utilisation non contrôlée des résultats obtenus. Nous nous aidons d'un étiqueteur automatique, TreeTagger et d'un logiciel de Textométrie, le Trameur. Les données sont suffisamment nombreuses (300 000 mots au moment où l'étude a été réalisée) pour pouvoir envisager les valeurs spécifiques de chaque futur, examiner les

cas où les deux formes sont en concurrence et voir si les corpus oraux apportent des arguments en faveur des théories qui s'affrontent sur l'étude de ces temps. Un autre intérêt du corpus est qu'il est assez homogène en ce qui concerne le genre pratiqué (des entretiens dont la plupart sont des trilogues favorisant des échanges détendus), cependant que varient le niveau d'études, le métier, l'âge, etc. des locuteurs. Il permet donc une première approche de la variation sociolinguistique à l'œuvre chez des Parisiens natifs d'origine.

Notre communication porte sur les usages de deux tiroirs verbaux<sup>1</sup>, le futur simple (FS) et le futur périphrastique (FP) dans un corpus d'interviews réalisées à Paris auprès de locuteurs d'âge et d'origine sociale diverses (CFPP2000)<sup>2</sup>. Il s'agit pour nous de valoriser les ressources fournies par le corpus CFPP2000, c'est-à-dire (i) de montrer que l'on peut apprendre des choses intéressantes sur le fonctionnement des deux formes FS et FP en partant d'activités langagières observables et de faits linguistiques attestés – et pas seulement des jugements d'acceptabilité introspectifs et d'exemples "cruciaux" forgés par des linguistes. (ii) de montrer l'aide apportée par les applications informatiques et statistiques qui permettent d'interroger ce corpus. Nous nous appuyons sur un outil statistique, le Trameur, et sur des concordanciers, qui sont eux aussi en ligne.

A chaque étape, nous envisageons les résultats apportés par ces outils qui permettent un parcours exhaustif et systématique du corpus et qui facilitent les croisements de données pour faire apparaître d'éventuelles corrélations entre des propriétés. Cependant nous insistons sur les précautions à prendre lorsqu'on pratique des analyses assistées par ordinateur. Nous décrivons le va-et-vient constant qu'il faut pratiquer entre les listes ou les décomptes fournis par les outils et le retour aux données de départ.

---

<sup>1</sup> Le terme de « tiroir » introduit par Damourette et Pichon (1911-1946) pour désigner les paradigmes de la conjugaison permet d'opposer les formes et leur valeur temporelle.

<sup>2</sup> Ce corpus en ligne à l'adresse <http://ed268.univ-paris3.fr/CFPP2000> (date accessed : 24/04/2010) se présente sous une forme alignant texte son et transcription. Il a été réalisé par une équipe comprenant S. Branca-Rosoff, S. Fleury, F. Lefeuvre et M. Pires.

## 1. Présentation du corpus CFPP2000

L'étude repose sur des données recueillies par entretiens non directifs auprès de 31 Parisiens, locuteurs natifs du français, un locuteur par entretien au moins habitant depuis sa tendre enfance dans le même quartier de Paris ou de la proche banlieue. Les métadonnées disponibles sur le site de CFPP2000 comportent des renseignements sociologiques (*cf.* figure1) qui permettent d'envisager de façon exploratoire des études contrastant l'âge ou la catégorie socioprofessionnelle des locuteurs (CSP). Comme nous le montrons pour les résultats résumés par la figure 13, ces catégorisations *a priori* peuvent – et doivent – être confrontées à ce que les locuteurs disent d'eux-mêmes et de leur trajectoire dans des entretiens longs (une heure en moyenne) qui les poussent à une activité réflexive intense, ce qui permet de confirmer ou de nuancer les CSP.

Figure 1 : exemple de descripteurs du corpus CFPP2000

Identifiant	Quartier	Age	Sexe	Situation familiale	Scolarité	Activité
YVETTE_AUDIN_F_70_7E	7	70	F	Veuve d'un avoca 3 enfants	Licence	a travaillé dans l'édition (retraîtée)
PAULINE_DE_BORDES_F_67_7E	7	67	F	Veuve d'un éditeu un enfant	Baccalauréat	Éditrice, directrice de galerie, co- fondatrice des ateliers des métiers d'art de la ville de Paris (ADAC)
RAPHAEL_LARIVIERE_H_23_7E	7	20	M	Père chirurgien ; mère infirmière 3 sœurs, une plus âgée est ingénieure, une	en 3e année de médecine	Interne

				de 17 ans entre en médecine ; une de douze ans		
MATHILDE_LELONG_F_85	11	88	F	Veuve (mari instituteur) trois enfants	Certificat d'études, puis CAP de couture	Mère au foyer
BLANCHE_DUCHEMIN_F_25_11 <sup>E</sup>	11	25	F	Célibataire Enfant unique Père formateur ; mère chercheur en sciences sociales (service de la recherche Institut National de l'audiovisuel) Tous deux sont de jeunes retraités	3e cycle en médecine en cours	Interne en médecine
THERESE_LE_VERN_F_70_	12	60	F	Divorcée	Certificat d'études primaires	Retraitée (Agent d'entretien)
NICOLE_NOROY_F_53_14E	14	53	F	Célibataire Parents commerçants (poissonniers)	Bac +3 licence	Libraire
ANDRE_MORANGE_H_58_MO	Montreuil	58	M	Marié, deux enfants	Certificat d'études primaires	Informaticien dans une banque

Le corpus, qui a continué à évoluer, s'approche des dimensions des corpus français déjà utilisés pour des études quantitatives (350 000 occurrences, au moment où nous rédigeons cet article). Il faut certainement considérer avec précaution les résultats obtenus, les caractéristiques idiolectales jouant un rôle dans la variation constatée (*cf.* l'exemple de *dire* en 5.1). L'impact du genre discursif doit être également pris en compte. Certains entretiens se rapprochent du style des conversations entre amis, tandis que d'autres s'apparentent à des entretiens formels, mais tous ont en commun d'être des paroles échangées à

*propos* des activités des sujets qui sont donc poussés à un certain décentrement par rapport au moment de l'énonciation, et non des paroles échangées *pendant* l'activité ce qui les aurait conduit à faire davantage référence à des actions en lien avec le moment de l'énonciation. Cette pratique a sans doute favorisé le FS au détriment du FP. Ainsi nos résultats ne valent pas pour « le » français tout court, mais ils donnent des indications sur les rapports de deux temps verbaux dans un *usage* de la langue. Nous pensons ainsi contribuer modestement à la construction d'une grammaire des usages du français. L'article est essentiellement consacré aux difficultés que rencontrent les approches quantitatives. Pour autant, dès qu'un corpus dépasse les 350 000 occurrences, le recours à des phases d'exploration informatique est une nécessité et garantit l'exhaustivité et la stabilité des résultats en fonction des critères adoptés. L'apport de l'informatique est aussi indispensable si l'on veut pouvoir relier entre elles les informations collectées.

## **2. Hypothèses**

S'agissant du futur, deux hypothèses sont en concurrence. Elles nous ont servi à choisir des indicateurs susceptibles de les départager.

### **2.1 Changement en cours...**

Selon la première, le FS va céder la place au FP. En l'absence de corpus comparables étalés dans le temps, et en suivant les propositions du linguiste Labov (1976 : 234 et s.), il est possible de raisonner en temps apparent : si un changement est en cours, la forme susceptible de l'emporter (le FP) doit être plus fréquente chez les jeunes, et la forme en recul, devenue un marqueur du style

soigné sera davantage attestée chez ceux que nous appelons des professionnels de la langue.

Pour démontrer cette hypothèse, il faut identifier puis compter les futurs pour chacun des groupes visés, ramenés à quatre pour les besoins de la démonstration : jeunes gens/personnes âgées ; professionnels de la langue/non professionnels.

## **2.1 ou bien valeur spécifique de chaque tiroir verbal**

Une deuxième hypothèse, soutenue par de nombreux chercheurs qui travaillent sur corpus, est que le FS se maintient bien et que les deux temps se répartissent les emplois<sup>3</sup>. Les uns comme Jeanjean (1988) et Bilger (2001) défendent la thèse d'une différence temporelle et aspectuelle avec le FS. Le FP est composé d'un présent, temps ancré dans l'actualité du locuteur et d'un infinitif indiquant le terme du procès (qui vient en quelque sorte borner le procès, comme le complément locatif vient borner *aller* dans son emploi de verbe de mouvement (« je vais à la gare » est parallèle à « je vais partir »). Le FS est ultérieur par rapport au présent, et rien ne vient le borner.

D'autres chercheurs, comme Laurendeau (2000) qui a travaillé sur des corpus canadiens, soutiennent la thèse d'une différence modale entre les futurs. Dans une communication précédente (Branca-Rosoff, à par.), l'une de nous avait relevé « à la main » pour une sous-partie du corpus, les occurrences des deux formes et fait une première estimation de l'influence des marqueurs temporels, aspectuels et modaux sur le choix du FS du FP. L'enquête montrait que le FP était susceptible de remplir tous les emplois temporels et que le futur simple dominait de façon

---

<sup>3</sup> Il faudrait également faire intervenir le présent à valeur de futur qui n'a pas été pris en compte dans cette étude (voir Le Goffic & Lab 2001 et Berit-Hansen et Strudsholm 2006).

nette quand il s'agissait d'emplois modaux et particulièrement dans un contexte négatif.

### **3. Préparation de la base textométrique BASE CFPP2000-Futur**

Le travail a porté sur l'ensemble des transcriptions « concaténées » et mises au format « *Lexico/Le Trameur* ». Le fichier de départ a été chargé dans Le Trameur (avec étiquetage morpho-syntaxique via TreeTagger<sup>4</sup>) ce qui a permis de construire une base lexicométrique annotée comportant 232 occurrences de FS et 333 occurrences de FP (après corrections).

Pour établir les données nous sommes partis des étiquetages automatiques opérés par l'étiqueteur TreeTagger (intégré au logiciel Le Trameur). Nous avons ensuite vérifié les étiquetages un à un, car les outils employés sont toujours susceptibles d'introduire des biais inaperçus au moment de leur construction. Loin de supprimer le travail de lecture les listes informatiques obligent à retourner au corpus pour vérifier en contexte que les étiquetages sont corrects. Cet établissement du corpus de travail, qui évoque l'établissement des textes anciens, est, si l'on veut, le premier pas vers cette activité de philologie numérique évoquée par Rastier (2001).

---

<sup>4</sup> <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/> (date accessed : 24/04/2010)

Figure 2 : Les catégories<sup>5</sup> (tree-tagger) dans la base CFPP2000-Futur

Fq	Catégorie (tree-tagger)▲	Num	
1	@card@	3	
135	ABR	3	
17837	ADJ	3	
40416	ADV	3	
33569	APOS	3	
475438	DELIM	3	
18281	DET_ART	3	
2700	DET_POS	3	
333	FP	3	Futur périphrastique
16	FP VER_mouvement	3	F. périphrastique ou V. Mouvement
232	FS	3	Futur simple
27	FS VER_cond	3	Futur Simple ou Conditionnel
18906	INT	3	
20680	KON	3	
10049	NAM	3	
35193	NOM	3	
4076	NUM	3	
19297	PRO	3	
7198	PRO_DEM	3	
1516	PRO_IND	3	
9	PRO_ON_GEN	3	
26978	PRO_PER	3	
23	PRO_PER4	3	
39	PRO_POS	3	
4677	PRO_REL	3	
27957	PRP	3	
7807	PRP_det	3	
42	PUN_cit	3	
1	SENT	3	
4918	UNKNOWN	3	
712	VER_cond	3	
7747	VER_impf	3	
5959	VER_infi	3	
18	VER_mouvement	3	V. Mouvement
5484	VER_pper	3	
480	VER_ppre	3	
38013	VER_pres	3	
782	VER_simp	3	
2	VER_subi	3	
501	VER_subp	3	

Nous évoquons rapidement les types de problèmes rencontrés.

<sup>5</sup> Après corrections sur les éléments visés par notre étude.



### 3.1 Des bévues de l'étiquetage

Nous avons découvert que les abréviations (en particulier les lettres composant des sigles comme H.L.M.) étaient automatiquement étiquetées comme des racines verbales ; nous avons dû corriger les confusions qui naissent des terminaisons des mots interprétées comme des suffixes du FS, par exemple (*des*) *mouчерons*, (*et*) *cetera* étaient traités comme des futurs<sup>6</sup>.

Il serait sans doute possible de proposer des corrections semi automatiques (par ex. en posant des règles sensibles au contexte prenant en compte les déterminants) et d'améliorer les outils. Ce travail n'a pas été mené. Nous retenons de cette étape qu'il faut toujours articuler des phases automatiques et des phases de contrôle où un utilisateur expert corrige les erreurs : travail indispensable mais qui peut être long.

Notons aussi que nous pouvons corriger les formes *introduites* par erreur, mais qu'il reste des incertitudes sur des formes éventuellement *omises*.

### 3.2 Des problèmes complexes de catégorisation : FP ou verbe de mouvement

Pour pouvoir examiner les propriétés du FP, il faut d'abord séparer le verbe de mouvement *aller* et les cas où *aller* + infinitive est la périphrase grammaticalisée qui nous intéresse. M. Gross avait en 1968 proposé des critères permettant de trier

---

<sup>6</sup> Mayaffre (2005) estime que TreeTagger et Cordial affichent dans leur fonction de désambiguïsation des homographes un taux de réussite avoisinant 100 %, et que les erreurs qui subsistent ne sont pas susceptibles de fausser le traitement statistique opéré par la suite. Lorsqu'on ne vise pas de grandes tendances, mais des propriétés grammaticales fines, on ne peut se satisfaire de cette marge d'erreurs. Par ailleurs, le taux d'erreurs augmente manifestement quand on traite des corpus oraux.

les constructions. Quand *aller* a les propriétés d'un verbe de plein statut, il est susceptible d'être employé à tous les temps (*j'irai travailler ; je suis allée travailler*) ; il construit des arguments qui peuvent être représentés par le pronom locatif *y* (*j'y vais*) et il impose des restrictions de sélection sur son sujet qui doit être compatible avec un verbe de mouvement et sur l'infinitif. Tel est le cas pour 1) :

Micheline : moi *j'avais travailler l'dimanche matin à huit heures* + je quitte mon domicile + *j'peux te dire qu'il y a une dizaine de personnes qui baladent leur chien euh + qui font pipi tous les matins c'est mignon c'est calme voilà* (Micheline Rosier, F, 56ans, 12<sup>e</sup>, enseignante<sup>7</sup>)

Lorsqu'*aller* + infinitif est un FS, sa conjugaison est restreinte au présent ou à l'imparfait :

Reine : je venais de + de *on va dire* de banlieue (Reine Cerêt, F-60 ans, 11<sup>e</sup>, chercheur institut national de l'audiovisuel)

De fait, dans (2), *aller* est incompatible avec le futur ou avec un accompli (*\*on ira dire ; \*on est allé dire*) et il ne se combine pas avec un pronom locatif (*\*on y va*). Enfin, *aller* partie d'une périphrase n'impose pas de restriction de sélection : il se combine très bien avec des verbes d'état :

tu *vas être* là tout l' temps (Pauline de Bordes, F-65 ans, 7<sup>e</sup> puis 6<sup>e</sup> éditrice)

et avec des sujets non susceptibles de mouvement réels ou métaphoriques :

ça va faire un an (Anita Musso, F., 46, 11<sup>e</sup>, auxiliaire de vie scolaire)

On peut déduire de ces spécificités que l'on n'a plus affaire à un verbe de plein statut, mais à une périphrase en voie d'intégration très avancée dans le système

---

<sup>7</sup> Chaque locuteur est identifié par un pseudonyme.

verbal (Vetters et Liere, à par. ; Bres et Labeau, à par.). 34 *aller* verbe de mouvement ont été éliminés du total du comptage initial réalisé par l'étiqueteur.

Lorsque les deux interprétations sont possibles les formes ont été incluses dans le compte des futurs (cette double annotation est visible *supra* dans le tableau des catégories de la base à l'issue des corrections réalisées). Cela a concerné quelques formes, dont l'énoncé suivant fournit un exemple :

Reine : on va s' retrouver à l'Entrepôt (Reine Cerêt, F-60 ans, 11<sup>e</sup>, chercheur institut national de l'audiovisuel)

Enq : *on va voir* ("on se déplace pour voir des habits" ou bien "on verra" (F-60 ans)

### **3.3 Des difficultés pour trier les FS et les conditionnels. La solution du double étiquetage**

Pour ce qui concerne le futur simple, il n'est pas évident de le distinguer du conditionnel lorsqu'il s'agit de la 1<sup>ère</sup> personne :

Raphael : non c'est ça s'est fait je sais pas trop comment + mais par exemple euh *j'dirai* que du primaire j'ai dû garder + dix pour cent d'mes amis + (Raphaël Larivière, H, 23, 7<sup>e</sup>)

*J'dirai(s)* peut s'interpréter comme un conditionnel : « (s'il fallait trancher *je dirais*, mes parents diraient » ou bien comme un FS, « *je vais dire*, *on dira*, dix pour cent ». *Je vais dire* est bien attesté dans le corpus et d'ailleurs des FS non ambigus apparaissent à la 2<sup>e</sup> et à la 5<sup>e</sup> personne : *tu m' diras* ; *vous m' direz*. Nous nous sommes heurtés à l'esprit binaire des classifications et des étiquetages grammaticaux (« c'est ou ce n'est pas un conditionnel ») et nous avons décidé de contourner le choix en proposant un double étiquetage : 59 sur les 215 FS comptés initialement par l'étiqueteur sont des *je dirai/dirais*, soit presque ¼ du

total. Ces formes ambiguës ont été prises en compte dans le premier comptage des FS, mais elles sont traitées à part par la suite. Ce problème de neutralisation se pose dans toutes sortes de zones problématiques du français<sup>8</sup>.

Voici un récapitulatif des formes :

Figure 2 : Les Futurs dans la base CFPP2000-Futur

Fq	Catégorie (tree-tagger) ▲	Num	
333	FP	3	<b>Futur périphrastique</b>
16	FP VER mouvement	3	<b>F. périphrastique ou V. Mouvement</b>
232	FS	3	<b>Futur simple</b>
27	FS VER cond	3	<b>Futur Simple ou Conditionnel</b>
18	VER mouvement	3	<b>V. Mouvement</b>

#### 4. L'outil informatique trameur

Les données traitées pour cette étude ont été soumises au logiciel Le Trameur<sup>9</sup>, développé par Serge Fleury (CLA2T/SYLED). Ce logiciel implémente une modélisation des ressources textométriques sous la forme d'une trame et d'un cadre textométriques (Söze-Duval, 2008).

La notion de trame textométrique permet de dépasser la segmentation initiale du texte en formes graphiques (mots) induisant toutes les annotations (de typage et de segmentation : propriétés, partitions), il est potentiellement possible de définir une trame suffisamment fine, et sans contenu propre (donc éventuellement modifiable) à laquelle rapporter toutes les informations (les

<sup>8</sup> Il ne s'agit pas de dire qu'il n'y a plus d'opposition paradigmatique entre les deux tiroirs, puisque les personnes 2, 3, 4, 4, 6 sont différentes. Cependant la première personne a des propriétés spécifiques et les *je dirai(s)* sont des formules d'introduction qui modalisent l'assertion.

<sup>9</sup> <http://tal.univ-paris3.fr/trameur/> (date accessed : 30/07/2010)

différentes dimensions), y compris donc la segmentation en formes graphiques. La trame se définit donc essentiellement comme une suite de positions.

Dans ce modèle textométrique, la partie **trame** de la ressource textuelle produite par le trameur est constituée par la suite des items isolés lors de l'opération de segmentation ; la partie **cadre** rassemble les données relatives aux différents découpages réalisés sur le corpus ainsi que les annotations projetées sur les contenants ou les contenus.

En résumé, un texte analysé dans Le Trameur est assimilable à une séquence ordonnée d'items (une liste de positions), chacun de ces items étant associé à une liste d'annotation paramétrable (forme, lemme, catégorie etc.). Ces items sont aussi regroupés dans des sélections de positions correspondant aux parties définies sur la trame (partie  $P_i$  de la partition  $PA_j$  allant de la position  $p_n$  à la position  $p_m$ ).

Les données analysées dans cette étude nous ont conduit à constituer une base textométrique (disponible sur le site de CFPP2000<sup>10</sup>). Dans cette base, la trame est constituée d'une liste d'items correspondant aux formes graphiques des contenus textuels des transcriptions, chacune de ces formes étant associée à son lemme et à sa catégorie ; le cadre permet de découper les contenus textuels en parties permettant d'isoler par exemple les différents entretiens, les tours de paroles etc.

Dans les parties suivantes, nous allons explorer la trame (dit plus simplement le contexte) des formes pivots que sont le FS et le FP.

---

<sup>10</sup> <http://ed268.univ-paris3.fr/syled/ressources/Corpus-Parole-Paris-PIII/work/BASE-cfpp2000-v11-11.zip> (date accessed : 24/04/2010)

#### **4.1 Examen de la base par le prisme FS versus FP via le Trameur**

Les fonctionnalités mises en œuvre dans Le Trameur permettent d'explorer la base constituée suivant différents points de vue.

Le concordancier permet par exemple d'explorer la trame textuelle au regard de l'annotation "catégorie", on peut donc visualiser des concordances des catégories visées par notre étude (les différents futurs).

Le calcul des cooccurrences permet de mettre au jour des proximités autour d'un pôle donné (par exemple, le pôle FP ou le pôle FS).

L'extraction de patron permet d'extraire sur la trame des séquences de positions consécutives sur la base d'un patron prédéfini (par exemple 5 positions telles que la troisième est un futur) : ce type de d'extraction permet un repérage approximatif de structures syntaxiques sous-jacentes. Aucune annotation syntaxique de type syntagmatique n'étant disponible dans la base construite par le Trameur, l'idée est ici de « s'approcher de la syntaxe » en repérant des séquences de la trame textuelle contenant des catégories pertinentes permettant de mettre au jour les éventuels syntagmes visés.

Le principe mis en œuvre pour réaliser l'extraction de patrons est le suivant :

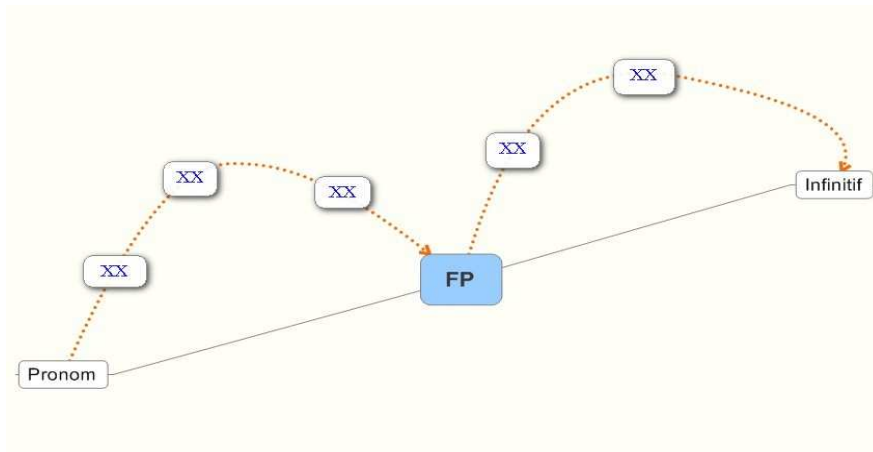
Afin de déterminer les éléments contextuels autour du pôle Futur (pronom, infinitif...), on extrait sur la trame une séquence de positions correspondant au modèle induit par le patron

La position des éléments contextuels visés pouvant varier à droite ou à gauche du pôle futur, il est nécessaire de procéder à plusieurs extractions pour construire la liste de tous les contextes attendus.

Si on considère le cas du FP, il est nécessaire de définir plusieurs patrons pour mettre au jour les différents éléments contextuels autour de ce pôle :

Figure 3 : Extraction progressive de patron

(FP+Infinitif) et (FP+XX+Infinitif) etc.  
(Pronom+FP) et (Pronom+XX+FP) etc.



Le travail d'extraction va donc se construire progressivement en essayant de mettre au jour par affinement progressif les séquences textuelles pertinentes induites par les patrons nécessaires et suffisants pour couvrir les contextes attendus.

## 5. Extraction de patrons

Si chaque forme verbale est encore proche de sa valeur d'origine, les deux futurs ne devraient pas être commutables dans de nombreux contextes. Si le FP a gagné

du terrain et s'apprête à remplacer le FS, il devrait se combiner avec les mêmes variables, les différences de fréquence étant plutôt interprétables en termes sociologiques. Pour confirmer ou infirmer l'hypothèse d'une différence de valeurs, on examinera les corrélations entre les tiroirs verbaux et les types de procès, la personne sujet, les adverbes de temps, les modalités.

### 5.1 Types de procès

Les types de procès exprimés par le lexique verbal peuvent être perfectifs (ou terminatifs, bornés) lorsque le contenu notionnel du verbe envisage le terme du procès : *mourir, trouver* ou *sortir* ou d'accomplissements : *traverser la rue, fumer une cigarette, tracer un cercle, etc.*

Ils peuvent être imperfectifs (non terminatifs, non bornés) quand le procès est envisagé dans son déroulement sans visée de terme. Tel est le cas d'états comme *être fatigué*, et d'activités comme *parler, marcher, chercher* qui peuvent se dérouler pendant un temps illimité. Le FS qui vise un point perspectif comme indiqué en section 1.2 devrait se combiner avec les perfectifs et le FS non borné avoir une affinité avec les imperfectifs. Pour voir si le choix des futurs est corrélé au lexique verbal on doit comparer l'infinitif qui suit l'auxiliaire *je vais travailler* avec le radical verbal du FS *je travaillerai*. Nous ne pouvons cependant nous borner à relever l'élément situé à droite d'*aller* car un ou deux éléments peuvent s'insérer, tels que la négation ou des arguments du verbe à l'infinitif :

j'    vais *pas* l' vouvoyer  
que l'on peut symboliser par :

auxFP+XX+Infinitif.



Il faut donc faire plusieurs tris : auxFP+ Infinitif ; auxFP + X+ Infinitif ; aux FP + XX + Infinitif. Nous nous bornerons ici aux résultats des verbes les plus fréquents qui permettent de donner des pourcentages.

Figure 4 : Les infinitifs du FP

	FS	FP	Total	FS %	FP %
DIRE	129	87	216	<b>59,7</b>	<b>40,3</b>
ETRE	29	31	60	<b>48,3</b>	<b>51,7</b>
AVOIR	33	12	45	<b>73,3</b>	<b>26,7</b>
ALLER	27	21	48	<b>56,3</b>	<b>43,8</b>
FALLOIR	8	3	11	<b>72,7</b>	<b>27,3</b>
FAIRE	14	29	43	<b>32,6</b>	<b>67,4</b>
VOIR	13	10	23	<b>56,5</b>	<b>43,5</b>
PARTIR	4	2	6	<b>66,7</b>	<b>33,3</b>
DONNER	3	3	6	<b>50,0</b>	<b>50,0</b>
POUVOIR	10	4	14	<b>71,4</b>	<b>28,6</b>

Il paraît impossible d'aboutir à des conclusions nettes en ce qui concerne l'opposition perfectif/imperfectif : ainsi, *avoir* (imperfectif) s'emploie bien davantage avec le FS, ce qui est conforme aux prévisions. Cependant *être*, verbe d'état, est équilibré : on trouve aussi bien dans le corpus, *vous serez médecin* que *vous allez être médecin, ça va être un métier stable*. Il paraît difficile de tirer des conclusions concernant la seule notion verbale. Peut-être le prédicat dans son ensemble, joue-t-il un rôle, mais les outils informatiques ne peuvent faire automatiquement ce calcul.

Les verbes modaux *pouvoir*<sup>11</sup> et *falloir* se combinent nettement mieux avec le FS. Deux explications peuvent être proposées. La première nous conduit vers l'hypothèse d'un FS incluant une part importante de modalité, alors que le FP serait davantage temporel. Nous verrons que la combinatoire avec les adverbes va aussi dans ce sens. On pourrait toutefois interpréter aussi ce fait en termes d'évolution. Parce que les modaux sont des éléments très fréquents ils résisteraient mieux aux évolutions. La prédominance de *pouvoir* et *falloir* + FS serait alors un archaïsme.

La forte proportion de *dire* au FS, tient en partie au hasard du choix des locuteurs dans un corpus encore trop restreint. *Dire* s'emploie très fréquemment comme une sorte de préfixe ou de suffixe d'énoncé avec une grande variation dans le choix des temps et des personnes « *je vais dire/je dirai/ on va dire/on dira* + proposition ». (« ça s' est beaucoup accentué moi j'dirai#FS|VER\_cond » (Pauline de Bordes, F, 67, 6<sup>e</sup>, éditrice ; « on va dire qu' j'ai des parents très très très sympa (Lucie da Silva, 22, 7<sup>e</sup>, employée de banque). Deux locuteurs sur la trentaine qui intervient dans ce corpus cumulent 49 des emplois de ce verbe au FS ce qui entraîne sans doute sa surreprésentation. Il aurait suffi que les locuteurs dont l'idiolecte privilégie « *je vais dire* » ou « *on va dire* » s'expriment davantage pour inverser les proportions.

---

<sup>11</sup> Il a été nécessaire de corriger les résultats fournis par TreeTagger sur un autre point. Le tableau indiquait tout d'abord qu'il y avait zéro occurrence de *pouvoir* au FP. Après vérification, grâce au concordancier, il est apparu qu'il s'agissait d'une anomalie d'étiquetage : le logiciel traitait toutes les occurrences de *pouvoir* comme des noms communs. L'exploration du corpus reste le meilleur moyen de repérer ces erreurs et de corriger manuellement les annotations proposées par le logiciel. Il est d'ailleurs possible avec TreeTagger de mettre en place un processus d'apprentissage de nouvelles connaissances à partir de données traitées. Cependant, dans notre étude, nous avons simplement réalisé l'opération de correction directement dans la base textométrique construite par Le Trameur.

## 5.2 La personne sujet

Si le FP fonctionne toujours comme un temps envisageant des évènements futurs à partir du moment de l'énonciation, on peut s'attendre à ce qu'il soit utilisé lorsque le point de vue subjectif est plus marqué que pour le FS. Plusieurs auteurs (Jeanjean 1988, Bilger 2001, Laurendeau 2000) ont d'ailleurs évoqué une forte affinité de FP avec la 1<sup>ère</sup> et la 2<sup>ème</sup> personne « j'crois qu ' j' vais reprendre un MAC » (Pauline 7e) et de FS avec la 3e personne : « ça disparaîtra pas » (Pauline, 7e).

Il faut donc repérer les syntagmes sujets, alors que – rappelons-le – il n'y a pas de repérage des fonctions dans le Trameur. Nous avons décidé de borner notre enquête aux pronoms, sans nous restreindre aux pronoms que TreeTagger étiquette comme personnels, ce qui exclurait les *ça*, *qui* etc. Il a fallu procéder à plusieurs extractions pour construire la liste de tous les contextes attendus.

On ne peut pas s'en tenir à un relevé des pronoms qui jouxtent la forme de futur, puisqu'il est possible d'insérer un *ne* discordancier et des clitiqes objets :

je n' le ferai pas  
je n'le lui donnerai pas

Nous avons donc fait là encore plusieurs tris : (Pronom + temps verbal) et (Pronom + XX + temps verbal), etc. Après quoi il a fallu trier à la main les résultats pour éliminer les pronoms toniques qui ne sont pas des sujets et qui apparaissent dans des énoncés comme « moi j'dirai(s) », ou plus gênant parce que la forme du pronom ne permet pas de repérer immédiatement l'erreur « Florence : pour vous ce sera # FS possible quand même de partir du quartier » (Corinne 7<sup>e</sup>). C'est aussi à la main qu'il faut voir si les *ça*, *qui*, *elle*, *nous* sont ou non des sujets : par exemple, *qui* n'entre pas dans le comptage dans une formule comme

« *qui tu connais* » ? *elle* dans « *elle je la rencontre tous les jours* », ou bien *nous* dans *il nous sera agréable*. On peut arriver ainsi au tableau suivant

Figure 5 : Les PRO et le FUTUR dans la base CFPP2000-Futur

	PRO FP				PRO FS		
	Occurrence	% FP	%(FP+FS)		Occurrence	%FS	%(FP+FS)
je	117	33,5	19,2		61	23,6	10,0
tu	9	2,6	1,5		11	4,2	1,8
il	9	2,6	1,5		10	3,9	1,6
elle	5	1,4	0,8		4	1,5	0,7
on	83	23,8	13,7		33	12,7	5,4
nous	0	0,0	0,0		2	0,8	0,3
vous	26	7,4	4,3		16	6,2	2,6
ils	25	7,2	4,1		7	2,7	1,2
elles	2	0,6	0,3		2	0,8	0,3
qui	15	4,3	2,5		6	2,3	1,0
ça	28	8,0	4,6		24	9,3	3,9
		91,4	52,5			68,0	28,9

%FP            proportion d'utilisation du pronom avec les FP  
 %FS            proportion d'utilisation du pronom avec les FS  
 %(FP+FS)    proportion d'utilisation du pronom avec les futurs FS et FP

On aura noté que les colonnes verticales ne sont pas sommées à 100% car 100% correspondraient à la somme de tous les pronoms y compris les objets ou les adverbiaux (*y*, *me*, *m'*, etc.) qui ont été retirés du tableau centré sur les pronoms sujets.

Les colonnes 2 et 5 s'expliquent par le genre discursif pratiqué par les interlocuteurs. Le sur-emploi de la 1<sup>ère</sup> personne correspond au fait que le thème des interviews poussent les enquêtés à parler d'eux et de leurs opinions, la présence important des *vous* au fait que la plupart des participants vouvoient l'enquêteur et vice-versa. Les résultats obtenus valent pour cette sphère d'application de la langue et d'autres situations entraîneraient d'autres usages des personnes. Les résultats confirment aussi le remplacement de *nous* par *on* en français contemporain.

Les colonnes 3 et 6 permettent de comparer pour chaque personne les différences d'utilisation des deux temps. On constate d'abord et surtout que le FP est plus disponible pour l'ensemble des usages. Cet écart se creuse pour les personnes 1 (19, 2% contre 10%) et pour *on* (13,7% contre 5, 4%). Les écarts baissent pour les autres personnes, ce qui suggère que le FS a davantage d'affinité avec la 3<sup>ème</sup> personne qu'avec les personnes renvoyant à la sphère de la subjectivité.

Ces résultats sont pourtant quelque peu à nuancer car la catégorie morphologique et l'effet discursif ne coïncident pas forcément. *On* a dans les données deux valeurs distinctes. L'équivalent d'un « nous » lorsque Marie évoque sa façon de faire les courses :

Marie : bah j ' vais faire les courses on va faire les courses à Carrefour  
(Marie-Louise Orsin, F. 64, 11<sup>e</sup>, institutrice)

Une valeur générique comme pour ces considérations sur la carte électorale parisienne :

à l'ouest on va avoir un peu plus de droite (Lucie da Silva, F, 22, 7<sup>e</sup>,  
employée de banque)

On ne peut pas davantage faire correspondre la catégorie morphologique du *tu* et une catégorie sémantique de 2<sup>e</sup> personne comme le montre l'énoncé suivant : « tu deviendras femme » est du DD où la locutrice se borne à relayer les sentences qui accompagnaient l'éducation sexuelle des jeunes filles :

Nicole Noroy : et comme maman a été réglée très jeune elle se doutait bien qu'y avait une des deux si c' n'est les deux à qui ça arriverait très jeunes + bon c'est tombé sur moi donc finalement + euh + elle a elle a j' pense que oui ma soeur devait avoir à peu près onze ans + et moi j'en avais neuf donc [...] donc X ça c'était pour ce qui était de de bon voilà du fait comme on disait "tu deviendras femme" < rire> + (Nicole Noroy, F, 53, 14<sup>e</sup>, libraire)

D'autres occurrences de deuxième personne s'interprètent de façon générique, se rapprochant de *on* et alternant d'ailleurs avec le présent :

Blanche : t'as d' l' argent pour boire plein d' verres tous les soirs mais t'auras jamais assez d'argent pour faire un resto (F, 25, 11<sup>e</sup>, interne en médecine).

### **5.3 Combinatoire avec des marqueurs temporels**

Beaucoup d'auteurs ont avancé que la virtualité du FS devait être compensée par des compléments de temps pour que cette forme puisse référer à une portion d'avenir précise ; au contraire, le FP, ancré dans la situation de parole, n'aurait pas besoin de délimiter le procès et ferait l'économie des adverbes de repérage temporel (Jeanjean 1988 : 241). Symétriquement, le FS qui ouvre un futur indéterminé aurait une forte affinité avec des adverbes ne localisant pas le procès.

Malheureusement TreeTagger ne dispose pas d'une catégorie adverbe de temps. Il faut donc trier à la main parmi les adverbes postposés au FS et au FP. Voici une des listes d'adverbes postposés au FS qui montre bien que la catégorie est en fourre-tout :

Figure 6 : ADV postposé au FS dans CFPP2000-Futur

ADV postpose au FS (juste à droite)

pas-ADV	47
plus-ADV	11
bien-ADV	6
jamais-ADV	5
toujours-ADV	3
ailleurs-ADV	3
vraiment-ADV	2
là-ADV	2
beaucoup-ADV	2
plutôt-ADV	2
ben-ADV	1
longtemps-ADV	1
effectivement-ADV	1
certainement-ADV	1
énormément-ADV	1
voilà-ADV	1
même-ADV	1
mieux-ADV	1

La recherche des marqueurs temporels qui intéresse le linguiste ne peut pas passer par le Trameur : les positions des adverbes sont trop mobiles. Nous avons donc eu recours à quelques sondages sur les cooccurrences.

Un premier tableau présente quelques sondages sur des marqueurs de localisation :

Figure 7 : Marqueurs de localisation

MARQUEURS DE LOCALISATION		
	FP	FS
<i>Repérage déictique</i>		
Maintenant	4	3
Prochain (mois, année, semaine...)	1	3
<i>Ponctuels non déictiques</i>		
A douze ans	(1)	
Subordonnées temporelles	9 et 3 <sup>12</sup>	10

A chaque étape, nous avons pu noter de nombreuses difficultés d'interprétation dont les plus importantes tournent autour du fait qu'un mot n'a pas une valeur sémantique fixe. Ainsi *maintenant* contrairement à ce qui est souvent écrit apparaît avec le FP. Cependant, il ne s'agit pas d'une datation : couplé avec le FS, *maintenant* signifie plutôt, comme *dorénavant*, le fait qu'à partir d'un point coïncidant avec le présent de l'énonciation débute un intervalle ouvert. A partir de marqueurs nous essayons de dire quelque chose sur un phénomène notionnel dans lequel la totalité de la proposition intervient. De même, couplée avec le FS la temporelle a une valeur de datation (« tu verras quand tu auras vingt ans »), alors que 9 des subordonnées au FP ont un effet itératif (« quand j'vais faire les courses » signifiant *chaque fois que*). Seules 3 subordonnées en *quand* localisent un évènement.

Lorsque la localisation du procès reste floue le FS est plus facilement sélectionné :

<sup>12</sup> Par exemple dans "j'vais être méchante quand j' vais dire ça +" l'énoncé semble renvoyer à l'évènement de parole qui en cours, (l'interview). Il s'oppose aux *quand* paraphrasables par "chaque fois que", qui ne devraient pas figurer dans le tableau et qui sont seulement mentionnées pour mémoire: seules 3 subordonnées en *quand* localisent un évènement.



Figure 8 : Localisation floue

LOCALISATION FLOUE		
	FP	FS
Un moment, un moment donné	2	4
un jour	2	1
fois	1	6

Il y a trop peu d'occurrences pour que l'on puisse raisonner sur les quantificateurs universels comme *toujours* (ou ses équivalents du type *toute ma vie*). *Jamais* se combine bien avec le FS et très mal avec le FP. Mais il s'agit d'un adverbe négatif. Nous y revenons au paragraphe suivant :

Figure 9 : Quantificateurs universels

QUANTIFICATEURS UNIVERSELS		
	FS	FP
Toujours	2	1
Toute ma vie	1	0
Jamais	5	0

Le FS a davantage d'affinité avec des localisations « floues » (8 occurrences sur 21 auxquelles on peut ajouter toutes sortes de cas non marqués (« comme on verra avec les enfants » au sens de « quand des enfants naîtront, on verra bien »).

#### **5.4 La combinatoire avec les adverbes négatifs**

Le FS est nettement associé à la négation. Il apparaît 66 fois contre 15 occurrences pour le FP.

j'ai fait ouais je lui dis + "non tu ne prendras pas de vélib +" (Laurence, F, 43 ans, 7<sup>e</sup>)

Yvette : on a dit "la rue d'Rennes n'deviendra **jamais piétonne**" (Yvette, F, 70 7<sup>e</sup>, correctrice dans une maison d'édition)

Figure 10 : Les Adverbes et les futurs dans la base CFPP2000-Futur

	ADV FP			ADV FS		
	Occurrence	% FP	%(FP+FS)	Occurrence	%FS	%(FP+FS)
pas	14	4,0	2,3	47	18,1	7,7
plus	0	0,0	0,0	11	4,2	1,8
jamais	0	0,0	0,0	5	1,9	0,8
toujours	1	0,3	0,2	3	1,2	0,5

Le FS cumule trois fois plus d'emplois que le FP.

## 6. Le versant sociolinguistique

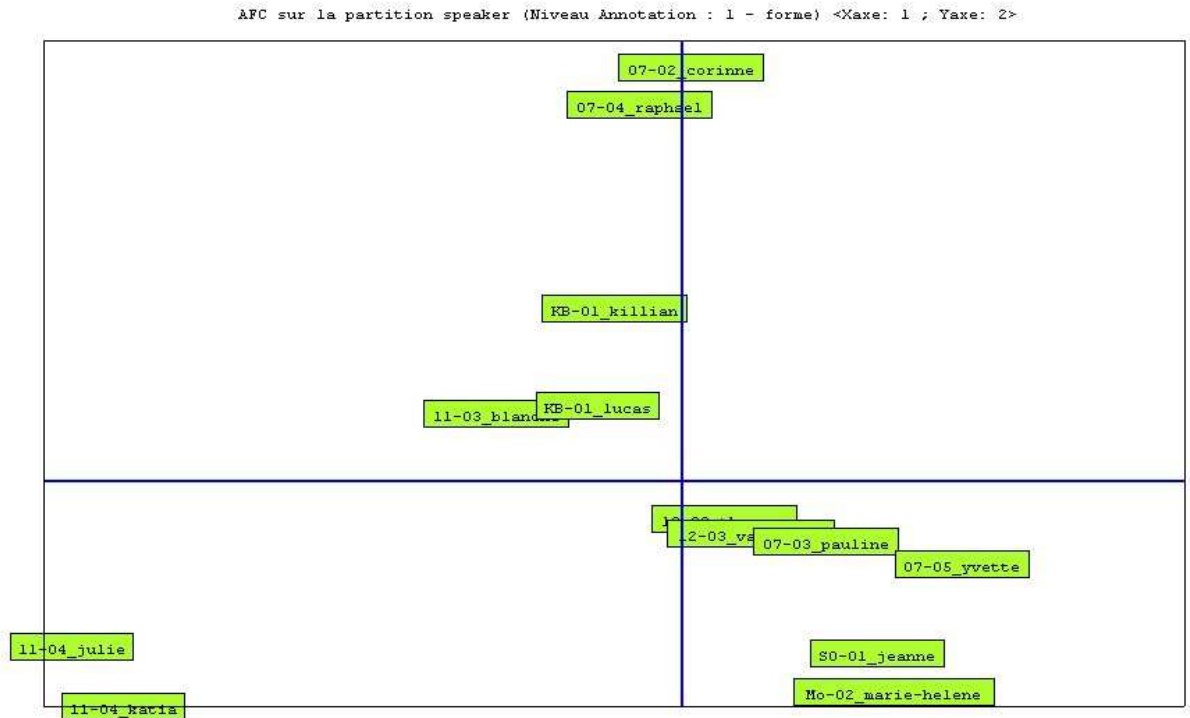
### 6.1 Un vocabulaire générationnel

Nous avons opposé 7 personnes de moins de trente ans et 7 de plus de 60 ans (en bleu sur le tableau ci-dessous). Nous examinons d'abord les résultats d'une AFC<sup>13</sup> sur l'ensemble du vocabulaire.

- AFC<sup>(1)</sup>(ensemble du vocabulaire) sur le corpus réduit aux groupes  
« Jeunes/Vieux »

<sup>13</sup> L'AFC (Analyse Factorielle des Correspondances) correspond à une famille de méthodes statistiques d'analyse multidimensionnelle, s'appliquant à des tableaux de nombres, qui visent à extraire des "facteurs" résumant approximativement par quelques séries de nombres l'ensemble des informations contenues dans le tableau de départ. Pour plus de détails, voir : <http://tal.univ-paris3.fr/wakka/wakka.php?wiki=glossairea>. Plus simplement et dans le cas traité ici : il s'agit de contraster les parties d'un corpus sur la base de leur vocabulaire respectif via un calcul statistique décrit plus haut : chaque partie étant une colonne du tableau et chaque ligne contenant un mot et ses fréquences dans les différentes parties, le tableau décrit donc la ventilation du vocabulaire sur chaque parti.)

Figure 11 : AFC "jeunes" vs "vieux" dans la base CFPP2000-Futur



Les résultats sont très nets : à l'extrême gauche du tableau on trouve deux lycéennes, nos locutrices les plus jeunes. En haut du tableau à gauche les étudiants et jeunes professionnels. A droite du tableau en bas, les locuteurs les plus âgés. Cependant, le futur joue-t-il un rôle de marqueur lié à l'âge ?

Figure 12 : Avant 30 ans versus après 60 ans

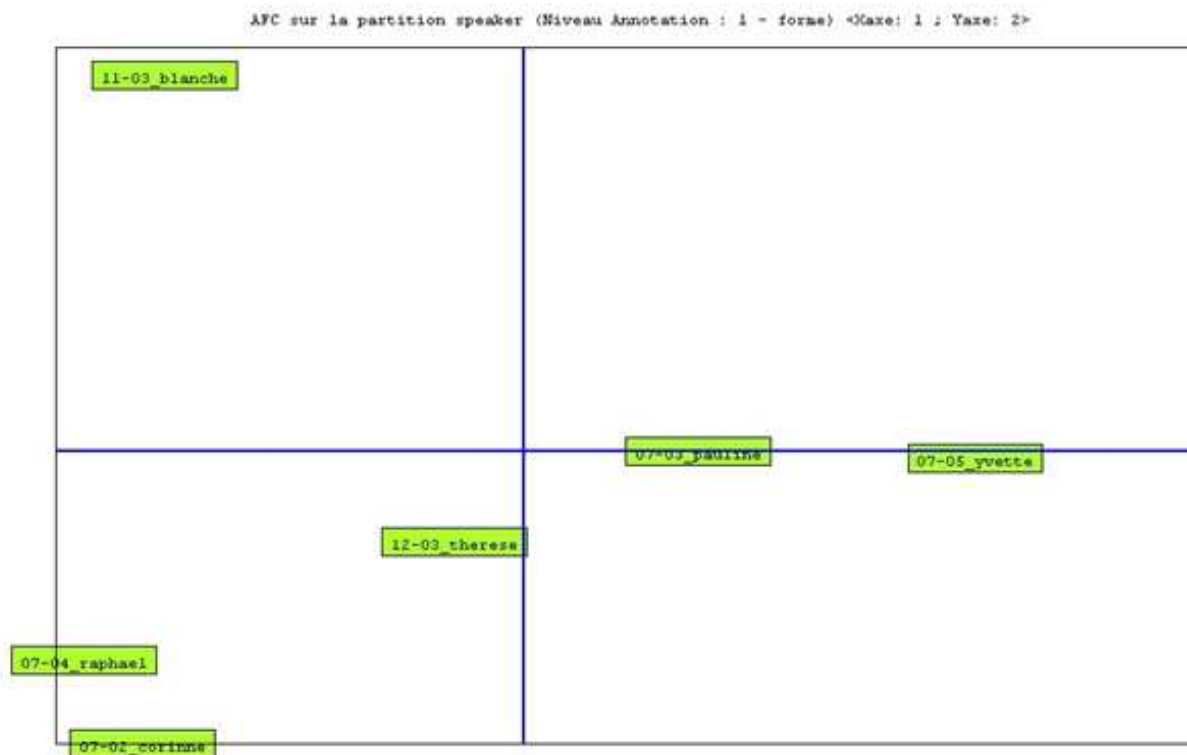
	KB-01_lucas	11-03_blanche	11-04_julie	KB-01_killian	07-02_corinne	11-04_katia	07-04_raphael		12-03_therese	12-03_valentine	07-05_yvette	SO-01_louise	SO-01_jeanne	Mo-02_marie-helene	07-03_pauline	
FP	8	22	0	2	36	7	13	88	1	18	15	3	2	7	8	54
FS	3	8	4	7	6	12	10	50	1	6	11	2	1	7	8	36
FS VER_cond	0	1	0	3	4	0	1	9	0	0	6	1	1	2	2	12

Il ne semble pas qu'il y ait eu d'effondrement et, petit fait significatif, Julie et Katia emploient davantage de FP que de FS.

## **6.2 Un vocabulaire corrélé à l'opposition professionnels/non professionnels**

Les critères sociaux sont très délicats à manier. Ainsi le baccalauréat des années 60, lorsqu'on entrait sur concours au lycée et que moins de 10% de la population avait ce diplôme a peu à voir avec le baccalauréat des années 2000. La grille des professions est assez opaque. Nous avons considéré cependant qu'il était intéressant de regrouper les locuteurs exerçant des fonctions en rapport avec la langue (éditeurs, enseignants, notamment) parce que ces derniers étaient particulièrement susceptibles de connaître les marqueurs du français standard. Ces « professionnels de la langue » ont été opposés à des locuteurs situés au pôle opposé parce qu'ils sont faiblement diplômés et qu'ils travaillent dans des métiers de service.

Figure 13 : AFC "pros" vs "pas pros" dans la base CFPP2000-Futur



L'AFC<sup>(1)</sup> pratiquée sur l'ensemble du vocabulaire du corpus aboutit à des contrastes intéressants de ce point de vue. A droite du corpus, on trouve deux locutrices âgées qui ont fait des carrières dans l'édition, Pauline et Yvette, qui s'opposent à Thérèse, gardienne d'école qui a commencé à travailler à 14 ans<sup>14</sup>.

Sur la partie gauche du tableau, on voit que la classe des locuteurs jeunes et diplômés est divisée en deux : Raphaël et Blanche sont deux étudiants en médecine, mais Blanche est isolée, loin des performances des professionnels de la langue. Un retour au corpus permet d'éclairer ce paradoxe : Blanche habite l'Est de Paris. Au cours de l'entretien, elle produit les catégories opposées de l'Ouest bourgeois et de l'Est « bobo ». Elle développe abondamment les caractéristiques

<sup>14</sup> On trouvera des métadonnées plus détaillées dans le tableau 1.

du style de vie bobo auquel elle s'identifie et il n'est donc pas étonnant qu'on puisse lui faire correspondre un style communicatif qui multiplie les marques du « parler jeune et branché ». L'AFC a donc permis un premier balayage du corpus et a attiré l'attention sur un écart de vocabulaire significatif : loin d'enfermer le chercheur dans des catégories pré-construites, la pratique statistique invite à confronter les résultats obtenues aux catégorisations des sujets.

### **6.3 Qu'en est-il des emplois du futur ?**

Même si l'échantillon est relativement faible, on peut calculer la répartition des futurs et leur spécificité sur ces 2 groupes, indice qu'il faudrait explorer plus précisément...

Figure 1

4 : Vocabulaire spécifique du groupe "jeunes" (exprimé en catégorie)

Forme	Ind-Specif	Eg-Totale	Eg-Partie
VER_impf	++	7747	1980
VER_pper	24.40	5484	1192
PRO_PER	12.40	26978	4861
PRP_det	7.31	7807	1455
KON	6.34	20680	3658
DET_POS	5.38	2700	525
PRO_REL	5.10	4677	872
DET_ART	4.25	18281	3189
FS VER_cond	4.20	27	12
VER mouvement	3.97	18	9
NUM	3.62	4076	743
UNKNOWN	1.96	4918	847
VER_simp	1.84	782	141
SENT	1.78	1	1
PRO_ON_GEN	1.75	9	3
PRO	1.67	19297	3237
PRO_IND	1.49	1516	258
PRO_PER4	1.48	23	5
NOM	1.36	35193	5841
ADJ	1.35	17837	2961
@card@	-1.08	1	0
VER_subi	-1.16	2	0
AER	-1.28	135	22
PRO_POS	-1.28	39	6
VER_subp	-1.47	501	79
FS	-1.51	232	35
VER_cond	-1.58	712	111
FP VER mouvement	-1.64	16	1
VER infi	-1.73	5959	961
FP	-1.89	333	47
VER_pres	-1.92	38013	6215
VER_ppre	-2.17	480	67
PRP	-2.23	27957	4537
APOS	-2.34	33569	5450
PUN_cit	-4.30	42	0
PRO_DEM	-13.26	7198	979
ADV	-14.11	40416	6194
NAM	-26.86	10049	1289
INT	-71.84	18906	2278

Figure 15 : Vocabulaire spécifique du groupe "âgé" (exprimé en catégorie)

Forme	Ind-Specif	Fq-Totale	Fq-Partie
PRO	**	19297	3363
APOS	25.76	33569	5230
ADV	12.03	40416	5969
VER_infi	11.34	5959	989
FP	9.83	333	87
VER_pres	7.87	38013	5543
KON	6.26	20680	3050
FS	4.66	232	52
PRO ON GEN	3.42	9	5
FS VER cond	3.11	27	9
PRO_IND	2.68	1516	236
VER_ppre	2.31	480	79
ABR	2.15	135	25
ADJ	1.93	17837	2500
@card@	1.86	1	1
VER_cond	1.65	712	105
VER_simp	1.54	782	113
PRO_PER4	1.41	23	4
SENT	-1.06	1	0
VER_subi	-1.13	2	0
FP VER mouvement	-1.21	16	2
VER_mouvement	-1.27	18	2
PRO_REL	-1.31	4677	640
VER_subp	-1.35	501	67
PRP	-1.38	27957	3823
UNKNOWN	-1.58	4918	659
DET_POS	-2.24	2700	342
PRO_POS	-2.64	39	1
PUN_cit	-3.69	42	0
PRO_DEM	-4.35	7198	892
NOM	-5.55	35193	4581
PRP_det	-5.94	7807	945
VER_pper	-6.32	5484	642
DET_ART	-8.06	18281	2273
NUM	-14.06	4076	406
INT	-18.61	18906	2208
PRO_PER	-23.35	26978	3182
NAM	-28.43	10049	1022
VER_impf	-40.59	7747	688

Les tableaux donnent des indications comptables sur les fréquences d'utilisation des catégories pour chaque groupe et du poids de celles-ci par rapport aux autres catégories : le FP et le FS sont plus spécifiques chez les jeunes (+9,83) que chez les personnes âgées (-1,89) et (1,51). Toutefois, on ne peut tirer des conclusions linguistiques de ce qui peut s'interpréter sur un plan thématique ou discursif : les jeunes évoqueraient davantage l'avenir que les personnes âgées, du moins dans le contexte particulier de l'interview.



Les spécificités relèvent plutôt d'une linguistique de la communication et le grammairien peut seulement constater que l'écart entre le FP et le FS (4.66) semble s'être creusé : 1.89-1.51 pour les personnes âgées = 0.38 et 9.83- 4.66 pour les jeunes = 5.17 pour les jeunes.

## 7. Conclusions

Nous nous gardons bien de tirer des conclusions définitives de cette brève exploration d'un corpus à l'aide de l'informatique. Constatons d'abord que le linguiste qui espèrerait se voir épargner le retour à l'observation des contextes aurait tort. Les résultats quantifiés imposent au contraire des retours constants à l'analyse fine des données pour corriger, visualiser les données traitées, réfléchir sur la valeur sémantique des formes. Les outils disposent d'ailleurs très souvent de fonctionnalités idoines pour permettre les retours au texte (le concordancier en est l'exemple le plus évident). D'autre part, les résultats proposés ne valent certainement pas pour *la langue française* ils constituent tout au plus des pistes pour une description des usages de la langue représentés dans un corpus de 350 000 mots dont la représentativité est donc encore un peu hasardeuse. Nous avons pu noter ainsi que le facteur du style (ou si l'on préfère de l'idiolecte) intervient manifestement dans les routines associées au verbe *dire*.

Mêmes imparfaits, les repères que nous avons adoptés permettent d'avancer trois idées.

Le FP ne fonctionne plus comme un "présent" suivi d'un ajout infinitif ce qui l'amène à remplir des fonctions qui vont bien au-delà de l'imminence. Il domine nettement dans les situations de localisation temporelle.

Le FS remplit sans problème des emplois temporels, mais il est moins fréquent que le FP dans cet usage.

Le FS écrase le FP dans certains emplois : il est particulièrement fréquent avec les verbes modaux, et avec des formes non assertives<sup>15</sup> en particulier les adverbes de négation. Il se combine bien avec des marqueurs temporels non localisants (comme *une fois*). Ces faits sont stables : on peut penser que le lien fort avec les emplois modalisants constitue un ensemble notionnel suffisamment systématique et cohérent pour qu'un FS peu assertif continue longtemps à fonctionner dans la langue française.

---

<sup>15</sup> Les exemples avec *si* sont peu nombreux, mais ils iraient dans le même sens

## Références

- Berit Hansen, A. & Strudsholm, E. (2006). 'Morphological and periphrastic future in French and Italian spoken language : Parallel tendencies ?'. In : Andersen, H. L., Birkelund, M. ; Mosegaard Hansen, M-B. (dir.). *La linguistique au cœur. Valence verbale, grammaticalisation et corpus : mélanges offerts à Lene Schøsler à l'occasion de son 60e anniversaire*. Odense : Syddansk Universitetsforlag : 189-218.
- Bilger, M. (2001). 'Retour sur le futur dans les corpus de français parlé'. *RSFP* 16, Université de Provence : 177-189.
- Branca-Rosoff, S. (à paraître). 'Corpus et recherches linguistiques sur le futur simple et le futur périphrastique'. In: Association for French Language Studies AFLS 2009, 3-5 September 2009, Université de Neuchâtel, Berne : Peter Lang.
- Bres, J. et Labeau, E. (à paraître). 'De l'espace au temps: des verbes de mouvements aller et venir à leur grammaticalisation en auxiliaires temporels dans quelques langues romanes (français, catalan, espagnol, italien, portugais) et en anglais'. In: Association for French Language Studies AFLS 2009, 3-5 September 2009, Université de Neuchâtel, Berne : Peter Lang.
- Damourette, E. et Pichon, J. (1911-1946). *Des mots à la pensée Essai de grammaire de la langue française*. Paris: d'Artrey.
- Gross, M. (1968). *Grammaire transformationnelle du français, Vol. 1, Syntaxe du verbe*. Paris : Larousse [réédité sous le même titre, Paris : Cantilène, 1986].
- Jeanjean, C. (1988). 'Le futur simple et le futur périphrastique en français parlé'. Blanche-Benveniste , A. Chervel & M. Gross (dirs.),. *Grammaire et histoire de la grammaire: Hommage à la mémoire de Jean Stéfanini*. Aix-en-Provence, Publications de l'Université de Provence : 235-257.

- Labov, W. (1976). *Sociolinguistique*. Paris: Minuit.
- Laurendeau, P. (2000). 'L'alternance futur simple / futur périphrastique : une hypothèse modale', *Verbum* XXII.3 : 277-292.
- Le Goffic, P. & Lab, F. (2001). 'Le présent "pro futuro"', *Cahiers Chronos* 7, Amsterdam-Atlanta, Rodopi.
- Mayaffre, D. (2005). 'Philologie et/ou herméneutique numérique : nouveaux concepts pour de nouvelles pratiques ?'. [http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Actes\\_ALBI-06.pdf](http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Actes_ALBI-06.pdf) (2007).
- Rastier, F. (2001). *Arts et sciences du texte*. Paris : PUF.
- Söße-Duval, K. (2008). 'Pour une textométrie opérationnelle' Article disponible en ligne : <http://tal.univ-paris3.fr/trameur/RTI6provisoire.doc> (date accessed : 30/07/2010).
- Vetters, C., Lière, A. (2009). 'Quand une périphrase devient temps verbal : le cas d'aller + infinitif'. *Faits de langue*, 33 : 27-36.