



HAL
open science

Références aux personnages dans L'occupation des sols : apport de la linguistique outillée

Frédéric Landragin, Noalig Tanguy, Michel Charolles

► To cite this version:

Frédéric Landragin, Noalig Tanguy, Michel Charolles. Références aux personnages dans L'occupation des sols : apport de la linguistique outillée. *Revue Sciences/Lettres*, 2015, Beaucoup de sens en si peu de mots. *L'Occupation des sols* de Jean Echenoz, 3, pp.2-21. 10.4000/rsl.816 . halshs-01153986

HAL Id: halshs-01153986

<https://shs.hal.science/halshs-01153986>

Submitted on 20 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Références aux personnages dans *L'occupation des sols* : apport de la linguistique outillée

Frédéric Landragin, Noalig Tanguy & Michel Charolles
LATTICE-CNRS, ENS & Université Sorbonne Nouvelle

DRAFT AUTEURS

Introduction

Le récit comprend quinze paragraphes et porte sur trois personnages humains principaux, le père, la mère et le fils Fabre. Le texte indique le prénom de la mère (Sylvie) et du fils (Paul), mais jamais celui du père, que l'on désignera donc par le seul nom de famille. A ce petit cercle familial, clairement défini dès la première phrase du récit, s'ajoutent divers personnages humains secondaires, plus ou moins définis (l'artiste Flers, son équipe, Jacqueline) ou indéfinis (quelqu'un, l'usager, un contremaître, des peintres, des hommes casqués de jaune). Le texte implique ainsi une quinzaine de référents humains. L'un des personnages principaux, Sylvie Fabre, a ceci de particulier d'être lié à une effigie, une peinture sur la façade d'un immeuble. Cette effigie n'est pas un référent humain en soi, mais l'intrication fréquente avec l'humain représenté va nous amener à l'ajouter à la liste des personnages étudiés. D'autres référents non humains tels que l'immeuble, l'espace vert à son pied ou le canal un peu plus loin constituent autant de « personnages » du récit, mais dans cette étude nous nous limiterons aux humains.

En complément de (Charolles, ce volume), nous allons étudier les caractéristiques linguistiques des mentions aux personnages, et comment celles-ci se répartissent selon les paragraphes du récit. Une mention telle que « Sylvie Fabre » ou « le fils Paul » est une expression référentielle, c'est-à-dire un ensemble de mots consécutifs du texte, qui permet d'identifier et donne accès à un référent, et permet ainsi à la phrase ou à la proposition courante de dire quelque chose de ce référent. Il s'agit d'une « forme explicite » de référence. Nous nous intéresserons aux types d'expressions référentielles, ainsi qu'aux phénomènes linguistiques qui ne constituent pas d'expressions référentielles mais qui contribuent néanmoins à évoquer un référent. Par exemple, dans « déménager, courir se refaire dans les grandes surfaces », l'utilisation de verbes à l'infinitif évoque les référents dont il est question, à savoir le père et le fils Fabre. Il ne s'agit pas d'expression référentielle ni même de mention, mais de ce que nous appelons une « forme atténuée » de référence, forme portée dans cet exemple par le verbe. Nous nous intéresserons aux rôles relatifs des formes explicites et des formes atténuées, de manière à couvrir l'ensemble des phénomènes référentiels du texte, du moins pour ce qui concerne les référents humains. Pour ce faire, nous utiliserons notamment la notion de « densité référentielle », définie ici comme le ratio entre le nombre de formes (explicites) de référence et le nombre de mots, à l'échelle du texte ou d'un paragraphe. Nous utiliserons également la notion de « chaîne de coréférence », en tant qu'ensemble ordonné des formes de référence relatives à un référent donné. Pour la détermination des données, les calculs et les analyses que nous serons amenés à faire, nous présenterons notre méthodologie ainsi que les outils utilisés. Extraire des données linguistiques d'un texte et analyser de manière qualitative et quantitative de telles données nécessite en effet des outils, et c'est ainsi que nous sommes amenés à parler de linguistique outillée. C'est avec la présentation de cette démarche que nous commençons cet article.

1. Méthodologie : annotation manuelle des formes de référence

Etudier un texte passe généralement par la formulation d'observations sur des éléments de ce texte, par exemple des marqueurs linguistiques. Ces observations peuvent inclure une ou plusieurs analyses linguistiques, en suivant un cadre théorique donné ou quelques règles spécifiées pour l'occasion. Elles constituent des annotations, qui viennent s'ajouter au texte pour en faciliter l'étude, celle-ci relevant alors de la linguistique de corpus (Habert *et al.*, 1997). L'annotation peut être effectuée manuellement ou automatiquement, ce qui requiert dans ce second cas l'exploitation d'outils de traitement automatique des langues, par exemple, pour ce qui concerne l'objet de cet article, d'identification des entités nommées, de résolution des anaphores et d'identification des coréférences. L'état actuel de ces outils n'est pas suffisant pour obtenir des analyses satisfaisantes : les phénomènes linguistiques les plus subtils sont ignorés et le taux d'erreurs est trop important pour pouvoir s'affranchir d'une étape de correction manuelle. Dans notre cas, la taille modeste du texte et le faible nombre de personnages s'ajoutent à ces arguments et nous conduisent à procéder à une annotation totalement manuelle. Choisir cette démarche nous entraîne à suivre une méthodologie et à utiliser des outils d'annotation précis (Fort, 2012).

La méthodologie comprend plusieurs phases. Une première phase consiste en la définition de l'objet d'étude, dans notre cas les marqueurs de référence aux personnages humains. De manière classique, il s'agit des expressions référentielles telles que les noms propres (« Fabre »), les descriptions définies (« le fils Paul »), les pronoms personnels (« il »), les possessifs (« sa » dans « sa hanche »), à partir du moment où un référent précis peut être identifié et où ce référent est humain. Cette première phase se fonde avant tout sur les travaux linguistiques portant sur la référence (Charolles, 2002). Chaque étude met cependant en avant un ou plusieurs aspects particuliers du vaste problème de la référence, chaque texte peut présenter des spécificités référentielles, et des éléments méthodologiques spécifiques peuvent venir ainsi adapter une méthodologie générale consensuelle. La première phrase de *L'occupation des sols* montre un tel exemple de spécificité : dans « comme tout avait brûlé – la mère, les meubles et les photographies de la mère –, pour Fabre et le fils Paul c'était [...] », l'expression référentielle « tout » réfère à un ensemble comprenant des référents non humains et un référent humain. Faut-il la retenir comme marqueur de référence pour notre étude ? Une réponse généraliste serait plutôt négative, « tout » n'étant pas une référence au personnage de la mère, mais nous répondons plutôt par l'affirmative, dans la mesure où « la mère » s'interprète par rapport à ce « tout » et permet d'expliquer comment le référent est introduit. De même, le personnage de la mère a ceci de particulier d'être évoqué de manière concomitante avec son effigie. Cette dernière joue un rôle important dans l'histoire, au point que nous l'avons retenue comme personnage à part entière, distinct de celui de la mère (quitte à fusionner les deux *a posteriori*), alors même qu'il ne s'agit pas d'un référent humain. Par ailleurs, « Fabre et fils Paul » coordonne deux marqueurs de référence, un premier référant au père et un second référant au fils. Faut-il également retenir la coordination comme marqueur de référence ? Une réponse généraliste serait plutôt négative, car en tenant compte de tous les groupements ainsi opérés on multiplie les référents. Ici, nous répondons plutôt par l'affirmative, dans la mesure où le groupe formé par le père et le fils est important dans le récit et en constitue un « personnage » quasiment au même titre que le père ou le fils individuellement.

Autre phénomène mis en avant par cette première phrase, la présence d'indices qui, sans référer, rappellent l'existence d'un référent. Nous l'avons dit dans l'introduction, c'est le cas dans « déménager, courir se refaire dans les grandes surfaces », où les sujets zéro – non exprimés – des infinitifs évoquent le référent, à savoir le groupe formé par le père et le fils, et contribuent à la saillance de celui-ci. Notre méthodologie tient compte de ces indices, parmi lesquels on trouve également les sujets zéro des participes prädicatifs, le second type des participes – par opposition aux participes adjectivaux – décrit dans (Fuchs *et al.*, ce volume). Nous nous intéressons ainsi à la fois aux formes explicites de référence et aux formes atténuées, ces deux types constituant les deux catégories de « maillons » qui vont nous servir à construire puis à étudier les chaînes de coréférence.

La deuxième phase de notre méthodologie consiste à repérer à l'aide d'un outil adapté les formes explicites et les formes atténuées dans *L'occupation des sols*, afin de les annoter une à une en indiquant à chaque fois l'identité du référent et en définissant ainsi les chaînes de coréférence. Suivant cette méthode, nous ne spécifions pas d'antécédent particulier à un pronom anaphorique. Au contraire, dès lors qu'à un tel pronom est affecté un référent précis, le pronom est intégré de fait dans la chaîne de coréférence caractérisant le référent en question. Notre méthodologie met ainsi en avant les relations coréférentielles plutôt que les relations anaphoriques (Landragin, 2011). L'intérêt de ce choix est illustré par l'exemple « à son fils il démontra cela sur plan » (§13). Dans cette phrase, « il » réfère au père, sans ambiguïté. Notre méthodologie consiste donc à annoter le pronom en ajoutant l'identifiant du père (qu'il s'agisse d'un champ textuel comme « Fabre » ou d'un indice comme « 1 »), ce qui se fait très rapidement. Si nous avons rendu obligatoire l'identification d'un antécédent, il aurait fallu décider entre la mention précédente dans le texte, à savoir « son » de « son fils » qui se trouve juste devant « il », et la dernière mention explicite (ou « antécédent plein »), qu'il s'agisse d'un nom propre ou d'une description définie singularisante. Dans ce cas, il faut alors remonter le texte : « son », puis « il » de « il dormait contre le sourire », puis « ses » de « ses calculs », puis « le » de « on ne le dissuada pas franchement d'emménager tout de suite », puis « s' » de « Fabre s'était présenté le premier », et enfin « Fabre », premier mot du §13. En plus de la complexité de la procédure et des problèmes d'identification éventuels, cette solution ne nous semble pas forcément plausible d'un point de vue cognitif. Nous lui préférons donc la solution consistant à construire les chaînes de coréférence.

L'outil utilisé pour cette phase est Analec (*Analyse de l'écrit*, cf. Landragin *et al.*, 2012 ; Victorri, 2013). La procédure consiste à définir tout d'abord la structure des annotations – ou schéma d'annotation – qui détermine quelles vont être les unités porteuses d'annotations et les valeurs possibles, ainsi que le type de structure de données informatique choisie pour représenter une chaîne de coréférence, à savoir un « schéma » dans le sens de (Widlöcher & Mathet, 2009 ; Mathet & Widlöcher, 2013). Elle consiste ensuite à utiliser l'interface de l'outil pour parcourir le texte et ajouter au fur et à mesure les annotations. Au final, 168 formes explicites et 71 formes atténuées ont été identifiées, à propos de 17 référents dont 10 référents individuels parmi lesquels la mère, son effigie, le père, le fils, Flers, l'usager, Jacqueline, le contremaître et d'autres personnes indéfinies. Deux annotateurs ont effectué l'identification des formes de référence (163 formes explicites identifiées dans une première version du corpus, 168 dans la deuxième version) et l'annotation en référents, de manière à détecter les cas ambigus et à relire attentivement le texte en cas de désaccord. Ce total de 239 formes de référence peut paraître important. Il l'est suffisamment pour qu'un outil soit nécessaire à l'appréhension des données annotées. Il ne l'est cependant pas suffisamment pour pouvoir réaliser des calculs statistiques avec significativités et tests de

corrélations. Autrement dit, cette étude met surtout au clair des tendances. Nous y reviendrons lors des analyses.

La troisième phase de notre méthodologie a consisté à annoter le texte, c'est-à-dire à ajouter aux phrases et aux formes de référence elles-mêmes des données décrivant des aspects morphologiques, syntaxiques, sémantiques et discursifs. Cet enrichissement vise à recueillir le maximum de données linguistiques, de manière à rapprocher les phénomènes référentiels d'autres phénomènes tels que la fonction syntaxique, le rôle thématique ou encore la position – préfixe, noyau, suffixe – des formes de référence dans les phrases. Avec cette phase se termine le travail d'annotation de *L'occupation des sols*.

La quatrième et dernière phase consiste alors à visualiser et à interroger les données annotées, c'est-à-dire à les parcourir (comme le texte, et d'ailleurs souvent en parallèle avec le texte), à en extraire quelques-unes en retenant une catégorie particulière, par exemple uniquement les pronoms de troisième personne ou les déterminants démonstratifs, puis à représenter graphiquement la distribution ou les caractéristiques linguistiques des exemples ainsi identifiés. Divers outils sont exploités dans ce but.

2. Outils pour l'étude des références et des chaînes de coréférence

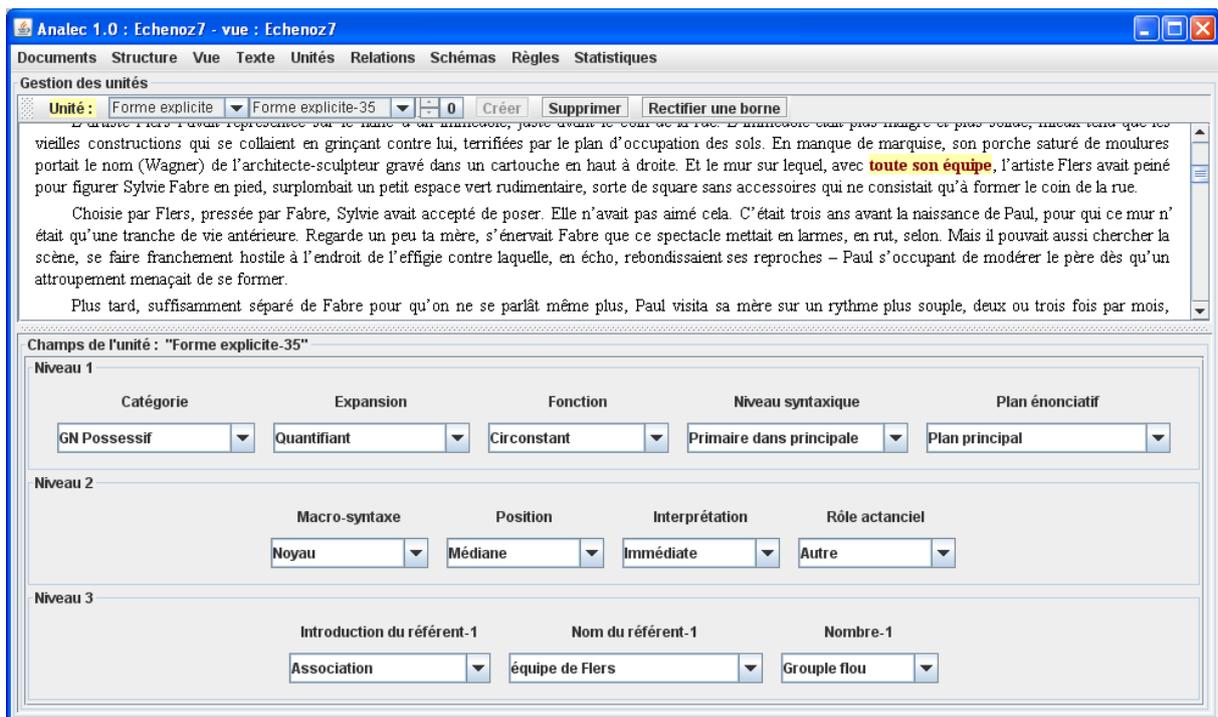


Figure 1 : Annotation avec Analec de l'expression référentielle « toute son équipe ».

Pour la phase d'annotation, on peut dans un premier temps utiliser les fonctionnalités de mise en relief (couleur de police, surbrillance du texte) d'un logiciel de traitement de texte. Rapidement, un outil dédié à l'annotation linguistique s'avère beaucoup plus efficace. Les outils disponibles sont nombreux et présentent tous des avantages spécifiques (Müller & Strube, 2001 ; Cunningham *et al.*, 2002 ; Morton & Lacivita, 2003 ; Habert, 2005 ; Avanzi *et*

al., 2008 ; Widlöcher & Mathet, 2009 ; Landragin *et al.*, 2012 ; Fort, 2012). L'important, c'est que l'annotation puisse se faire de manière ergonomique, rapide et efficace, et que les données annotées soient enregistrées dans un fichier informatique facilitant les exploitations ultérieures : visualisations graphiques variées, interrogations, constitution de base de données.

Nous avons importé le texte complet de *L'occupation des sols* dans le logiciel Analec (Victorri, 2013), et nous avons spécifié une structure d'annotation regroupant les unités « forme explicite » et « forme atténuée », ainsi que le schéma « chaîne de coréférence ». L'interface d'annotation d'Analec permet alors de délimiter les formes de référence, de les annoter et de les regrouper dans des chaînes de coréférence. C'est ce que montre la figure 1 : l'expression référentielle « toute son équipe » du §4 a été délimitée (le texte correspondant apparaît sur fond jaune), et vient d'être annotée en tant que « forme explicite ». Les différents types d'annotation apparaissent dans une interface à base de menus déroulants, en-dessous du texte. On y distingue trois niveaux d'annotation qui regroupent en une vue synthétique les données propres à la forme de référence, à la phrase et à la chaîne de coréférence.

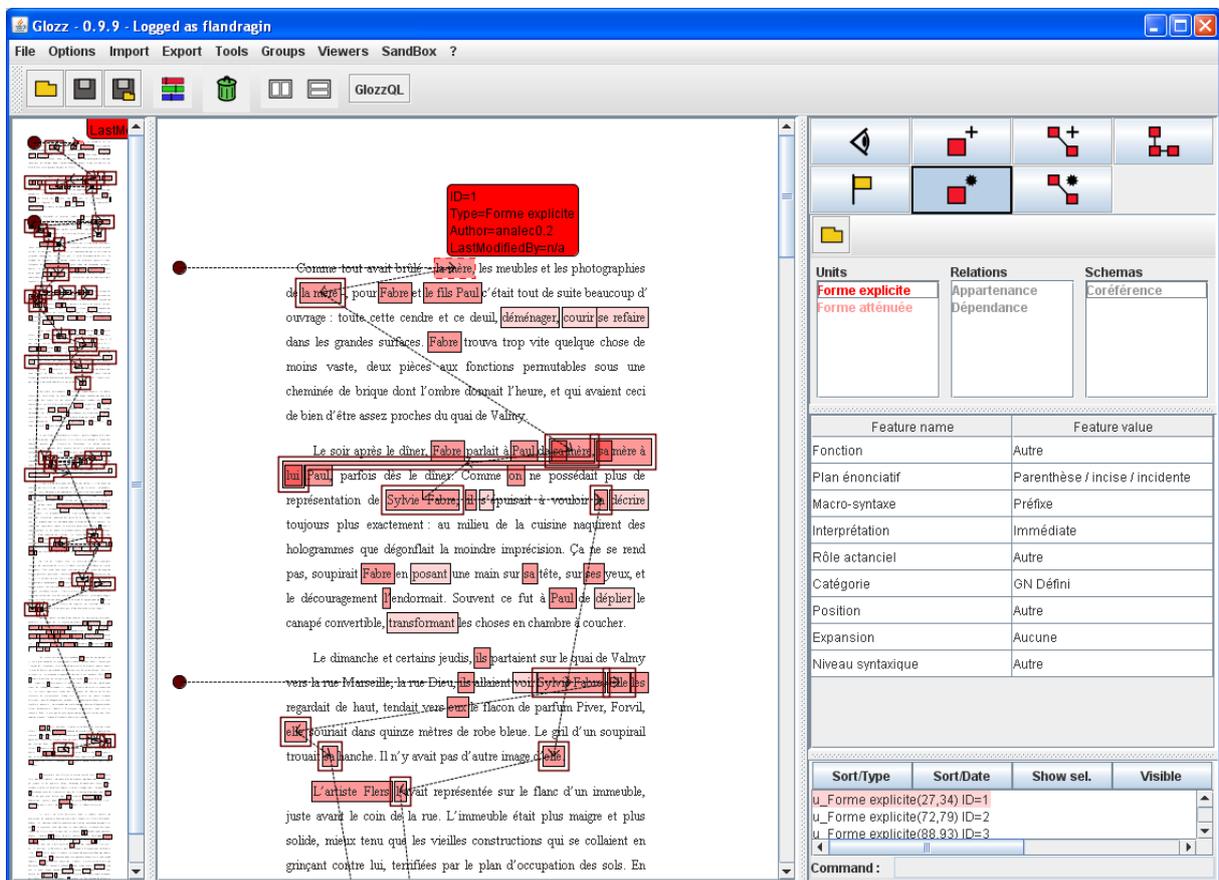


Figure 2 : Visualisation avec Glozz de l'ensemble des formes de référence annotées dans le texte.

Quand on s'intéresse aux chaînes de coréférence d'un texte, il est utile de pouvoir visualiser la répartition des maillons sur l'ensemble du texte, personnage par personnage ou toutes à la fois. C'est ce qu'illustre la figure 2 issue de l'outil Glozz (Mathet & Widlöcher, 2013), outil dans lequel les annotations réalisées avec Analec ont été importées. Cette figure présente l'interface principale de Glozz, avec un premier volet vertical à gauche qui montre

l'étendue des chaînes de coréférence sur tout le texte (celui-ci est alors rapetissé de manière à en voir la totalité), un deuxième volet vertical qui montre la zone de travail avec les formes explicites et atténuées déjà repérées, et un troisième volet qui regroupe quelques boutons de contrôle de l'interface ainsi que l'ensemble des champs correspondant ici à l'expression « la mère » de la première phrase. Deux chaînes de coréférence sont mises en avant graphiquement : celle correspondant à la mère, qui commence à la première ligne, et celle correspondant à son effigie, qui commence au début du §3. Toutes les formes référentielles du texte apparaissent sur un fond rouge. Celles qui sont affichées avec un fond rouge sombre sont les formes explicites, celles qui sont affichées avec un fond rouge clair sont les formes atténuées.

Pour cette étude, nous avons utilisé à la fois les outils Analec et Glozz : chaque outil a ses propres avantages, et – à partir du moment où les formats d'enregistrement du corpus sont compatibles – le linguiste peut passer à loisir de l'un à l'autre. Parmi ses avantages, Glozz est doté de nombreuses fonctionnalités de visualisation et d'interrogation de corpus. De son côté, Analec est très souple, permet de modifier la structure des annotations en cours de travail, et peut faire apparaître à tout moment un « concordancier », c'est-à-dire une liste d'exemples alignés, liste constituée en fonction d'un critère de sélection. L'intérêt d'une telle liste est double : tout d'abord, en rapprochant dans un même espace visuel des exemples similaires, elle permet de les comparer et de les enregistrer pour se constituer un stock d'exemples remarquables ; d'autre part, elle permet d'identifier des erreurs d'annotation. En effet, un exemple qui n'a rien à voir avec les autres est probablement mal annoté (ou alors c'est un cas particulier, intéressant à ce titre). Analec permet de corriger directement l'erreur à partir du concordancier, si erreur il y a. Nous soulignons cet aspect dynamique de l'annotation car il nous semble essentiel en linguistique outillée. En effet, nous modifions souvent en cours de travail notre façon d'annoter, nos champs, nos valeurs, et les outils doivent prendre en compte dynamiquement ces besoins, plutôt que d'imposer une phase d'annotation strictement distincte de la phase de lecture des données. C'est ce qui s'est passé lors de cette étude. Lors du travail du premier annotateur, des questions se sont posées, notamment sur la frontière parfois floue entre forme explicite et forme atténuée : où classer les pronoms réflexifs ? Un premier choix a été fait, qui a entraîné la spécification d'une structure d'annotation impliquant certains pronoms réflexifs dans les formes explicites et d'autres dans les formes atténuées. Puis, compte tenu des difficultés de décision rencontrées, nous avons décidé de catégoriser tous les réflexifs en tant que formes explicites. Il n'a pas été nécessaire de reprendre le texte à partir du début : Analec s'est chargé de corriger dynamiquement les annotations avec la nouvelle structure. Par ailleurs, les interrogations des annotations auxquelles nous avons procédé ont parfois généré des tableaux de nombres et de fréquences d'occurrences. Un dernier outil que nous avons utilisé est ainsi un tableur, de manière à produire les graphiques des figures 4, 7, 8 et 9.

3. Analyse des formes de référence

Nous allons étudier comment réfèrent les mentions aux personnages humains et en déduire quelques observations sur les spécificités du texte. Une étude rapide des 168 formes explicites montre que le nom propre complet (prénom et nom) n'est utilisé que pour un seul personnage, Sylvie Fabre. Le père est désigné par le nom de famille seul, au point qu'on ne connaît même pas son prénom ; le fils est désigné par son prénom Paul ; les autres personnages sont désignés soit par le nom de famille seul (Flers, Wagner), soit par une description plutôt qu'un nom propre, avec éventuellement l'exception de Jacqueline, qui n'est

désignée par ce prénom que dans un dialogue, en tant que terme d'adresse, et qui est introduite dans le texte par la description pour le moins indéfinie « une femme ». Le nom propre a ceci de particulier de référer de manière directe et non ambiguë à son référent. En théorie, « Sylvie Fabre » réfère forcément à la mère, alors que « Fabre » peut référer au père, au fils ou à la mère. Sylvie Fabre constitue ainsi une sorte de pivot référentiel du récit. La présence et l'importance de son effigie vient bouleverser ce mécanisme : au §3, dans « ils allaient voir Sylvie Fabre », le nom propre complet réfère bien à l'effigie et non à la mère. Dans la suite, le texte comprend deux autres mentions « Sylvie Fabre » et deux mentions du prénom « Sylvie » qui réfèrent sans ambiguïté à l'effigie et non à la personne. En regardant les mentions à la personne, on trouve également deux fois le nom propre complet et deux fois le prénom seul. Pour ce qui concerne l'emploi du nom propre, tout est donc fait pour mélanger les références de la personne avec celles de son effigie. C'est une première spécificité de *L'occupation des sols* : la mère et son effigie occupent une place centrale, et, si l'on est bien en présence de deux référents, les expressions référentielles utilisées pour l'un ou l'autre sont quasiment identiques, incitant à ne pas les distinguer. Les seules différences entre les deux résident d'une part dans la description définie « la mère », qui ne vaut que pour le référent humain (alors que « ta mère » et « sa mère » sont utilisées pour les deux référents), et dans l'emploi des déterminants possessifs, qui apparaissent avec des éléments relatifs à la peinture (« sa hanche », « son parfum », « son effacement personnel », « ses deux dimensions », cf. figure 3), mais qui n'apparaissent pas pour le référent humain (cf. figure 4, en comparant la ligne dédiée au portrait de Sylvie Fabre avec celle, juste au-dessus, dédiée à Sylvie Fabre elle-même).

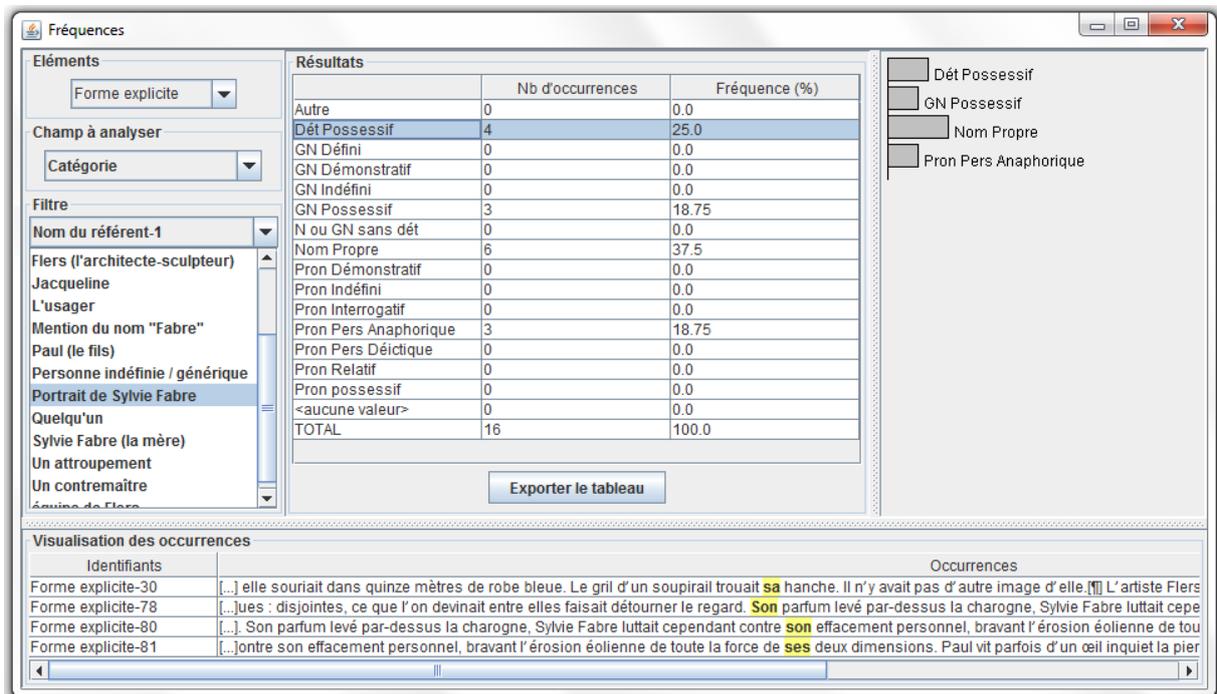


Figure 3 : Fréquences des catégories des formes explicites, ici pour l'effigie de la mère. On voit notamment que la catégorie la plus fréquente est le nom propre (37,5% des formes explicites) et que le texte comprend 4 exemples incluant un déterminant possessif, exemples qui sont affichés dans le concordancier de l'outil (Analec).

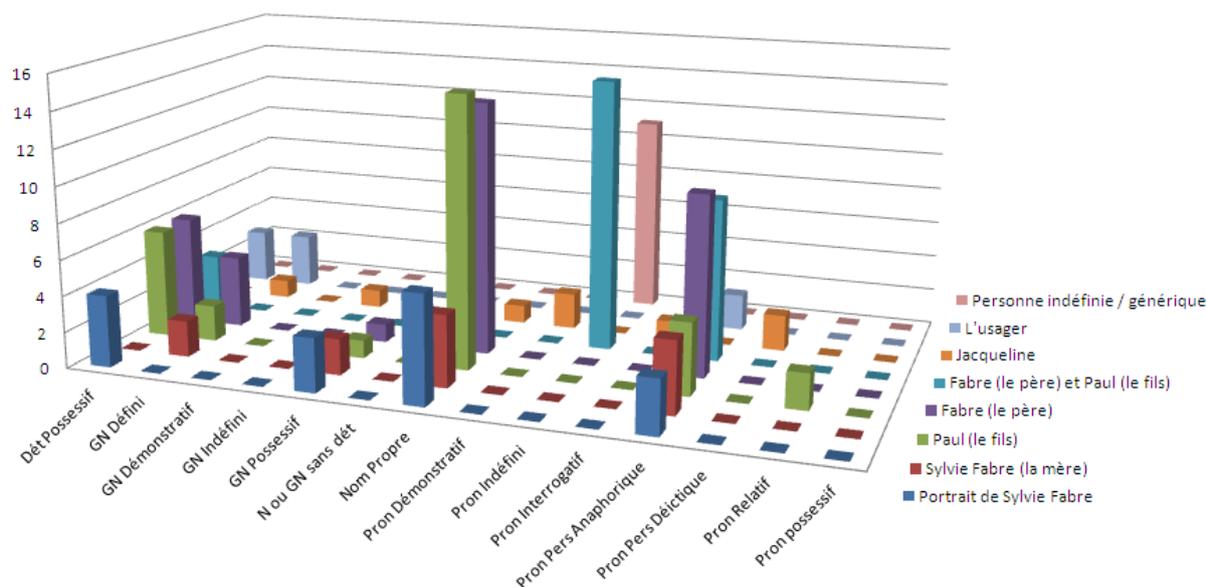


Figure 4 : Catégories des formes explicites pour chaque personnage, avec les nombres d'occurrences. On voit notamment que le nom propre n'est utilisé que pour les 4 personnages principaux (avec une exception pour « Jacqueline ») et que certaines catégories ne donnent lieu à aucune mention : GN démonstratif, pronom interrogatif, pronom possessif.

Parmi les référents humains du récit, seuls le père et le fils forment par moment un groupe, au sens d'un référent collectif désigné par des formes au pluriel : « ils », « eux », « leur ». Si l'on met de côté les expressions plurielles telles que « les peintres » et « des hommes calmement casqués de jaune », dont les référents ne sont jamais désignés sous une autre forme que ces descriptions, le groupe formé par le père et le fils est ainsi le seul référent pluriel (humain) du récit. Nous sommes donc en présence d'une part d'un double référent pivot relatif à la mère et d'autre part d'un groupe composé par le père et le fils. Ce groupe n'empêche pas les référents individuels Fabre et Paul de jouer leur propre rôle dans l'histoire, mais il ajoute une sorte de référent supplémentaire, essentiel dans l'histoire, sans hiérarchie entre les deux. (Charolles, ce volume) souligne que d'une manière générale, le père précède le fils au début du récit, cette tendance s'inversant à la fin. L'étude des mentions, par exemple des formes possessives « son père » ou « son fils » (les deux apparaissent à peu près aux mêmes endroits, de même que « le père de Paul », mais avec très peu d'occurrences), ne suffit pas à hiérarchiser les deux référents. L'étude des fonctions syntaxiques des mentions du père et du fils ne suffit pas non plus : les deux ont comme fonction la plus fréquente la fonction sujet, au début comme à la fin, dans 54% des cas pour le père et 30% pour le fils (la figure 5 montre un exemple de test de corrélation mis en œuvre pour ce type de calcul). L'étude des rôles actanciels montre que le père est le plus souvent agent (54% des cas), de même que le fils, dans une moindre mesure (40% des cas), mais les statistiques ne sont pas significatives et il est difficile d'en déduire que le père est plus mis en avant que le fils. Même remarque pour la macro-syntaxe : le père comme le fils sont majoritairement mentionnés en noyau (71% des cas pour le père, 73% pour le fils). Il nous reste donc à regarder les phrases qui regroupent père et fils sous un même prédicat. Au début du récit, on trouve ainsi « Fabre parlait à Paul » (§1), avec le père précédant le fils, puis l'inverse avec « Paul s'occupant de modérer le père » (§5). Après la période de séparation des deux personnages, on trouve « il avait failli appeler son père » (§6), « Paul aperçut Fabre » (§9), « à son fils il démontra » (§13), « Fabre était plus disert que Paul » (§14), autrement dit un mélange des deux ordonnancements. Il semble donc

que le père et le fils forment un groupe cohérent, avec deux éléments ayant à peu près autant d'influence l'un que l'autre. Nous noterons que sur les 168 formes explicites, 30 réfèrent à ce groupe, ce qui en fait un référent presque aussi important que le père (41 formes), que le fils (32 formes), et autant que la mère avec son effigie (30 formes), pivot de l'histoire par son rôle et par l'usage du nom propre, mais pas par la fréquence des références.

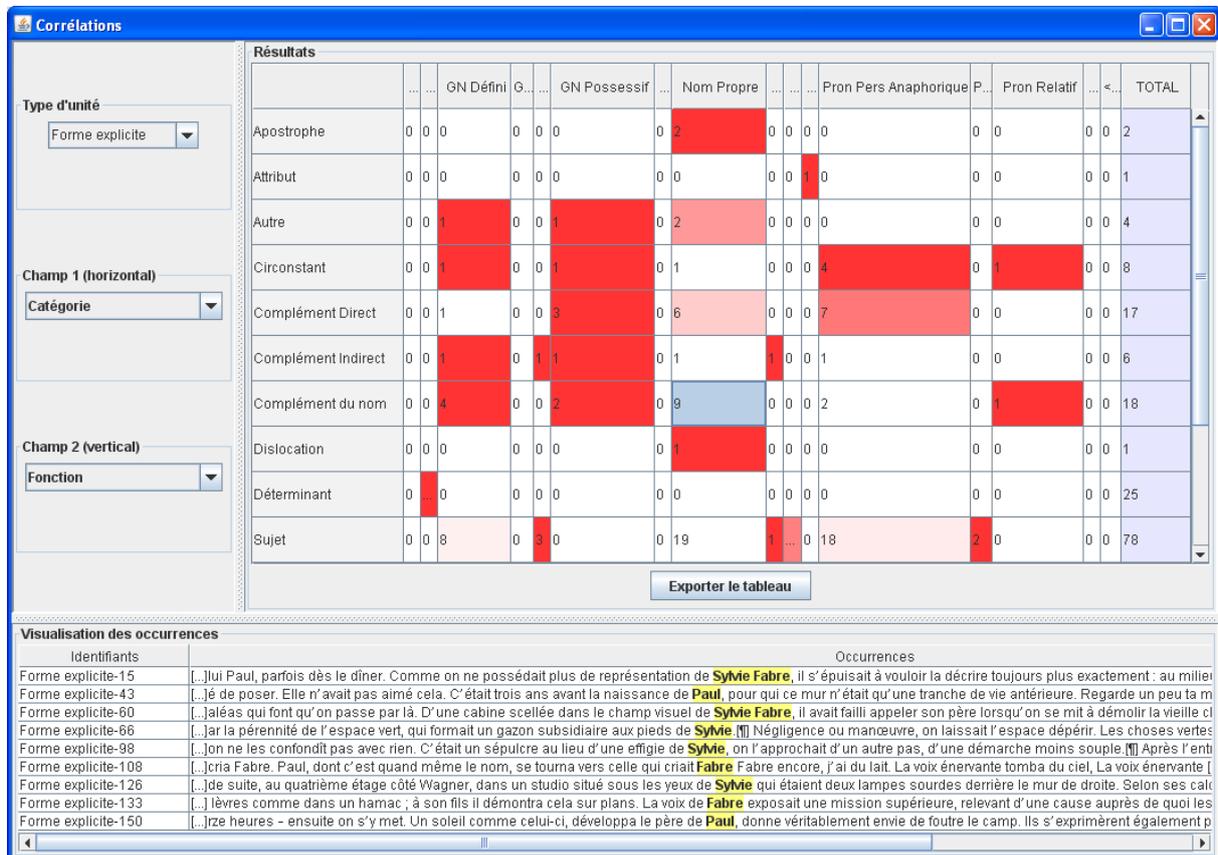


Figure 5 : Rapports entre les catégories et les fonctions syntaxiques des formes explicites. On voit notamment dans cette copie d'écran d'Analec que la principale catégorie ayant la fonction de complément du nom est celle du nom propre (cf. les exemples dans le concordancier). On y voit également que les pronoms anaphoriques « il » et « elle » ont majoritairement la fonction sujet.

Quels que soient les personnages, les formes de référence dans *L'occupation des sols* s'avèrent classiques (cf. figure 4) : nom propre, description définie (« la mère », « le père », « le fils Paul », « l'effigie », « l'artiste Flers », « l'usager »), description indéfinie (« une femme », « un attroupement »), pronom personnel (« il », « elle »), groupe nominal possessif, avec parfois quelques variantes jouant sur les relations entre les personnages : « sa mère, sa mère à lui Paul » (§2), « le père de Paul » (§15, alors que cela fait longtemps que le lecteur a compris qu'il n'y avait qu'un seul père dans toute l'histoire). Les adjectifs qualificatifs, les épithètes sont absents, de même que les démonstratifs : l'auteur n'utilise jamais ni « celui-ci », ni « celle-ci », ni même « ce N » avec un N qui pourrait être « père », « fils » ou « mère », mais qui pourrait aussi redénommer le référent. Le démonstratif marquant une proximité psychologique avec le personnage (Charolles, ce volume), il semble que l'auteur ait choisi une forme de neutralité dans les références à ses personnages, autre spécificité du texte qu'il fallait souligner. Les termes pleins qui apparaissent dans les formes explicites vont dans

ce sens : à part « mère » (8 occurrences, ce qui appuie le statut de pivot de ce personnage), « père » (4 occurrences) et « fils » (4 occurrences), on ne trouve qu'une seule occurrence d'un autre terme – « le veuf » (§9) – redénomination dont le registre est encore une fois celui de la famille.

La forme de référence la plus fréquente est le pronom, et notamment le pronom personnel indéfini : « on » (18 occurrences), « il » (9), « son » (8), « sa » (8), « ses » (5), « elle » (4), « ils » (4). Ce calcul de fréquences montre tout d'abord que les pronoms personnels « il » et « elle » ne sont pas plus fréquents que « le père » et « la mère », ce qui veut dire que les personnages principaux sont souvent désignés par leur rôle familial. Il montre ensuite la forte fréquence des références indéfinies, que l'on retrouve aussi avec « quelqu'un », « une femme » ou « un attroupement ». Or « on » a ceci de particulier qu'il peut référer aussi bien à un ensemble précis de personnages qu'à un référent générique, voire au narrateur et au lecteur. Dans cette étude de la référence, nous avons dû décider à chaque occurrence de « on » si le pronom référait à un groupe précis, notamment celui formé par le père et le fils, ou s'il pouvait référer à un ensemble beaucoup plus flou, incluant des membres dont il est impossible de dresser la liste. Au final, 15 occurrences sur les 18, donc quasiment toutes, réfèrent à un groupe incluant le père et le fils (cf. la barre la plus élevée dans la figure 4). C'est le cas dès le §2 avec « comme on ne possédait plus de représentation de Sylvie Fabre », et c'est bien entendu le cas de l'intégralité du dernier paragraphe, avec l'exemple « on gratte, on gratte et puis très vite on respire mal, on sue ». Comme il est difficile de se prononcer sur le caractère strict du groupe désigné ainsi par « on », nous considérons que le référent est ici un groupe flou, qui comprend de manière sûre le père et le fils, mais que l'on pourrait également étendre (de manière floue) au lecteur en train de suivre le récit (Charolles, ce volume ; Fuchs *et al*, ce volume). L'interprétation est en tout cas homogène, et très différente de celle que nous donnons à « on ne sait pas qui c'était » (§11, où « on » vaut quasiment pour Paul seul), ainsi que de la valeur générique que nous donnons à « on laissait l'espace dépérir » (§7), « ce que l'on devinait entre elles faisait détourner le regard » (§8), « on lança la superstructure » et « les vitres étaient encore barrées de blanc d'Espagne pour qu'on ne les confondît pas avec rien » (§10).

Dans tous les cas, et cette remarque est valable pour l'ensemble des formes de référence du texte, l'interprétation n'est pas ambiguë entre deux ou plusieurs choix possibles. Lors de toutes ses occurrences, le pronom « on » réfère sans ambiguïté, soit au groupe incluant le père et le fils, soit à Paul seul, soit à une personne générique. De même, toutes les mentions incluant « mère », « elle » ou « sa » réfèrent sans ambiguïté soit au personnage de la mère, soit à son effigie. C'est le cas même dans « sa mère, sa mère à lui Paul » où aucun des annotateurs et des lecteurs impliqués dans cette étude n'a imaginé que le texte pouvait parler de la mère du père de Paul. Autrement dit, il n'y a pas d'endroits du texte où se pose un problème durable de calcul de la référence, et l'annotation en références (à des humains) n'a donné lieu à aucune ambiguïté effective, ce qui, sans être spécifique à *L'occupation des sols*, n'est pas si fréquent dans les textes littéraires ou dans les textes de presse (Landragin, 2011).

4. Analyse des chaînes de coréférence

Nous allons maintenant étudier comment se structurent les chaînes de coréférence et en déduire quelques observations sur l'accès aux personnages dans le récit. Une chaîne de coréférence regroupe de manière ordonnée – dans l'ordre linéaire du texte – l'ensemble des mentions à un personnage donné. Comme le montre la figure 6, une chaîne de coréférence se

représente graphiquement par une liste chaînée de formes de référence. On constate par exemple que la chaîne de coréférence dédiée au personnage de la mère commence – outre l’expression « tout » non annotée ici – par « la mère », puis enchaîne avec les expressions « la mère », « sa mère », etc. Un peu plus loin dans le texte (§3), la chaîne de coréférence dédiée à l’effigie commence par le nom propre « Sylvie Fabre » et enchaîne avec le pronom anaphorique « elle ».

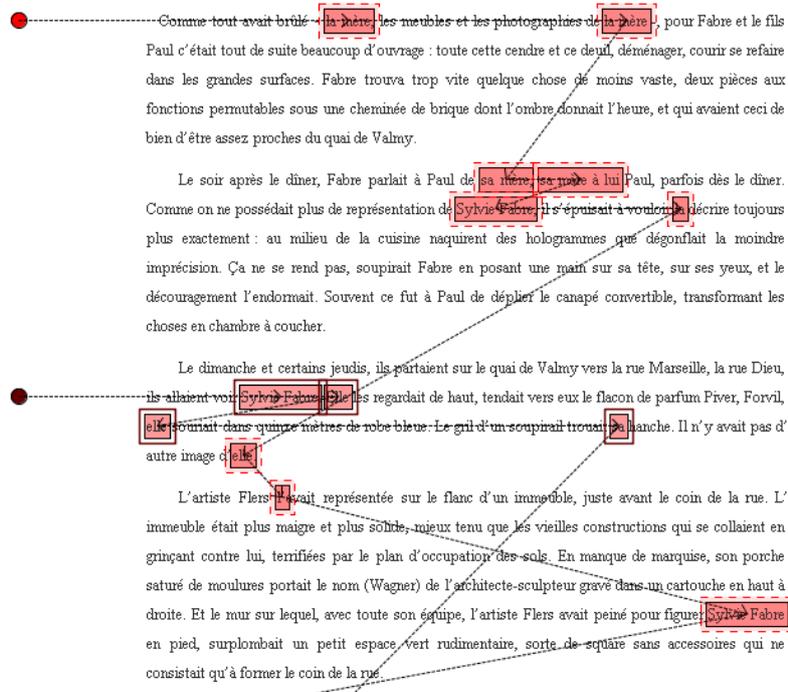


Figure 6 : Le début du texte avec les chaînes de coréférence correspondant à la mère (en rouge clair dans cette copie d’écran de l’outil Glozz) et à son effigie (en rouge foncé).

La liste complète des maillons des chaînes de coréférence, extraits du texte, est donnée dans la table 1. On y retrouve toutes les formes explicites, ainsi que le nombre de maillons de chaque chaîne de coréférence, c’est-à-dire le nombre de mentions pour chacun des personnages. Concernant la mère et son effigie, on remarque tout d’abord que le référent humain est extrait d’un « tout » pour être ensuite désigné plusieurs fois de suite par son rôle de mère. Le nom propre « Sylvie Fabre » n’apparaît ainsi qu’en tant que cinquième maillon (début du §2), ce qui est bien loin du prototype de chaîne de coréférence qu’est la suite : nom propre – description définie – succession de pronoms personnels de troisième personne. On remarque également que de son côté, l’effigie est introduite dans le texte par le nom propre « Sylvie Fabre », puis par quelques pronoms, et, en cinquième position dans la chaîne, par un groupe nominal avec « mère » comme tête. Ce début de chaîne est déjà plus proche d’une chaîne typique d’un référent humain, ce qui n’est pas sans conséquences. Nous l’avons déjà vu, les expressions référentielles utilisées pour la mère et son effigie sont les mêmes, ce que nous vérifions ici avec « Sylvie Fabre », « ta mère », « sa mère », « la mère » et « elle ». Avec en plus une chaîne structurée pour l’effigie comme pour un référent humain, nous constatons que tout est fait pour perturber le lecteur et lui faire prendre du non humain pour de l’humain. De fait, en première lecture, le premier maillon de la chaîne de l’effigie peut être facilement pris comme un maillon supplémentaire de la chaîne de la mère. Ce n’est qu’après avoir lu

« elle souriait dans quinze mètres de robe bleue » que les questions que l'on se pose (mais n'est-elle pas morte ?) trouvent une réponse et que peut se construire la chaîne de coréférence dédiée à l'effigie – éventuellement en revenant en arrière. Dans la suite du texte, les deux chaînes de coréférence s'entrecroisent, qui plus est avec des maillons identiques : « Sylvie », « Sylvie Fabre », « sa mère ». Cependant, une fois les référents clairement posés, et en deuxième lecture pour ce qui concerne le début du texte, aucune ambiguïté ne reste et c'est pourquoi les deux annotateurs ont bien identifié « Sylvie Fabre » comme le premier maillon d'une nouvelle chaîne.

Personnage	Chaîne de coréférence	Proportion de noms propres
Fabre, le père	Fabre – Fabre – Fabre – il – s' – Fabre – sa – ses – l' – Fabre – s' – Fabre – que – il – ses – le père – Fabre – son père – Fabre – le veuf – Fabre – se – il – Fabre – s' – le – ses – il – ses – son – il – Fabre – le père – Fabre – s' – il – se – lui-même – il – le père de Paul – Fabre	32% (13 sur 41)
Paul Fabre, le fils	le fils Paul – Paul – sa – sa – lui – Paul – Paul – Paul – qui – ta – Paul – Paul – sa – il – son – Paul – Paul – sa – se – Paul – il – Paul – lui – Paul – se – Paul – son fils – du fils – il – Paul – Paul – Paul	50% (16 sur 32)
Groupe formé par le père et le fils	se – on – ils – les – eux – on – se – on – leur – on – on – on – on – s' – on – s' – ils – s' – leur – ils – leurs – on – s' – se – on – on – on – on – on – on	0%
La mère	(tout) – la mère – la mère – sa mère – sa mère à lui – Sylvie Fabre – la – elle – l' – Sylvie Fabre – Sylvie – elle – Sylvie	31% (4 sur 13)
L'effigie de la mère	Sylvie Fabre – elle – elle – sa – ta mère – l'effigie – sa mère – Sylvie Fabre – Sylvie – son – Sylvie Fabre – son – ses – Sylvie – sa mère – l' – Sylvie	35% (6 sur 17)
Flers	l'artiste Flers – son – l'artiste Flers – Flers	75%
L'usager	l'usager – l'usager – s' – l'usager – sa – il – se – son – sa – soi	0%
Jacqueline	une femme – qui – s' – celle – qui – j' – tu – Jacqueline – la femme – s' – qui – c'	8%

Table 1 : Formes explicites des chaînes de coréférence. Les personnages qui n'apparaissent pas ici sont ceux qui ne créent pas de véritable chaîne de coréférence, c'est-à-dire ceux qui ne sont pas mentionnés avec suffisamment de formes explicites.

Contrairement à la mère, les personnages du père et de Paul sont introduits de manière claire et précise, le père par le nom propre « Fabre » qui est répété plusieurs fois de manière à bien ancrer le personnage, c'est-à-dire à bien signaler que le nom de famille utilisé seul désigne ce personnage, et le fils par l'expression sans mystère « le fils Paul » qui réfère autant qu'elle définit celui-ci. Les deux chaînes de coréférence sont ainsi construites autour des maillons identifiants que sont « Fabre » pour le père (32% des occurrences, soit plus que « il », « le père » et que toute autre forme de référence) et « Paul » pour le fils (50% des occurrences). La chaîne de coréférence du père commence naturellement par une alternance de « Fabre » et de pronoms. Celle du fils n'est quasiment, d'un bout à l'autre, qu'une alternance de « Paul » et de pronoms, ce qui en fait une chaîne très linéaire, sans surprise. Seules deux mentions du « fils » (§13) remettent en avant son rôle, au moment des retrouvailles, comme s'il fallait que tout soit clair lors de cette étape importante. Cette remise en avant est plus fréquente pour Fabre : comme le montre la table 1 avec les caractères gras, le mot « père » est utilisé à plusieurs reprises, alors qu'un simple pronom aurait parfois suffi. Par

exemple, « Paul s’occupant de le modérer » aurait pu remplacer « Paul s’occupant de modérer le père » (§5), de même que « il » ou « Fabre » aurait pu remplacer la redénomination « le veuf » (§9) et que « Fabre » aurait pu remplacer la description sur-spécifiée « le père de Paul » (§15). Sans refaire les références, il s’agit de souligner que plus que les autres personnages, le père est sujet à des redénominations et à un rappel de son rôle de père, alors même qu’il s’agit du personnage le plus mentionné dans le texte. Les interprétations sont immédiates : difficultés à surmonter la mort de sa femme, à garder son fils près de lui...

Un dernier point sur lequel (Charolles, ce volume) s’interrogeait est la reprise anaphorique, et notamment la nature de l’antécédent de « il » ou « elle ». Nous voyons dans la figure 5 que le pronom anaphorique est le plus souvent sujet (18 cas sur 32), puis complément d’objet direct (7 cas), puis circonstant (4 cas). Ce résultat ne dit cependant pas grand-chose sur la saillance du référent, et il nous faut tenter une corrélation non pas entre la catégorie et la fonction syntaxique des maillons, mais entre la catégorie d’un maillon de type anaphorique et la fonction de son antécédent. Comme le décrit notre procédure d’annotation dans la section qui lui est dédiée, nous n’avons pas accès facilement à un tel lien dans notre corpus, et il nous faut procéder à cette étude manuelle, au cas par cas. C’est l’objet de la table 2, qui montre rapidement que la tendance est celle d’un antécédent sujet, avec l’exception notable de l’effigie de la mère, à savoir le seul référent non humain de la liste. L’accès aux antécédents s’avère ainsi tout à fait classique dans *L’occupation des sols*.

Pronom	Antécédent(s)	Fonction de l’antécédent plein
« il », §2	« Fabre », phrase précédente	sujet
« il », §5	« Fabre », phrase précédente	sujet
« il », §12	« Fabre », phrase précédente	sujet
« il », §13	« ses » (même phrase), puis « le » (phrase précédente), puis « s’ » et enfin « Fabre » (phrase précédente et 1 ^{ère} phrase du §13)	sujet
« il », §13	« il » ci-dessus, ce qui fait remonter à « Fabre » de la 1 ^{ère} phrase du §13	sujet
« il », §14	« Fabre », proposition précédente	sujet
« il », §14	« il » ci-dessus, puis « Fabre » (phrase précédente)	sujet
« il », §6	« Paul », phrase précédente	sujet
« il », §10	« Paul », phrase précédente	sujet
« il », §14	« Paul », phrase précédente (§13)	sujet
« ils », §3	« on », §2, puis « Fabre et le fils Paul », §1	objet indirect
« ils », §15	difficile à déterminer (soit via « le père de Paul » dans la phrase précédente, soit « on » dans la phrase précédente)	sujet
« ils », §15	« leur », puis « ils » ci-dessus, avec le même problème d’identification d’un antécédent plein	sujet
« elle », §5	« Sylvie », phrase précédente	sujet
« elle », §3	« Sylvie Fabre », phrase précédente	objet direct
« elle », §3	« elle », proposition précédente (ci-dessus), puis « Sylvie Fabre », phrase précédente	objet direct
« il », §7	« sa » puis « l’usager », phrase précédente	sujet

Table 2 : Fonction syntaxique des antécédents des pronoms anaphoriques, pour les personnages suivants (dans l’ordre des lignes du tableau) : le père, le fils, le groupe formé par le père et le fils, la mère, son effigie, l’usager.

5. Analyse des rapports entre référents et paragraphes

Après les références et les chaînes de coréférence, nous allons maintenant étudier la densité référentielle et en déduire quelques observations sur la structure du récit. Comme décrit dans l'introduction, nous avons défini cette notion de densité référentielle comme le ratio entre le nombre de formes explicites et le nombre de mots. Comme nous l'avons fait tout au long de cet article, les référents sont les personnages humains auxquels s'ajoute l'effigie de la mère, et nous allons scinder le texte en paragraphes, afin de procéder aux calculs au niveau de chaque paragraphe et non au niveau du texte en entier. Tous référents confondus, le calcul de la densité référentielle conduit au résultat de la figure 7.

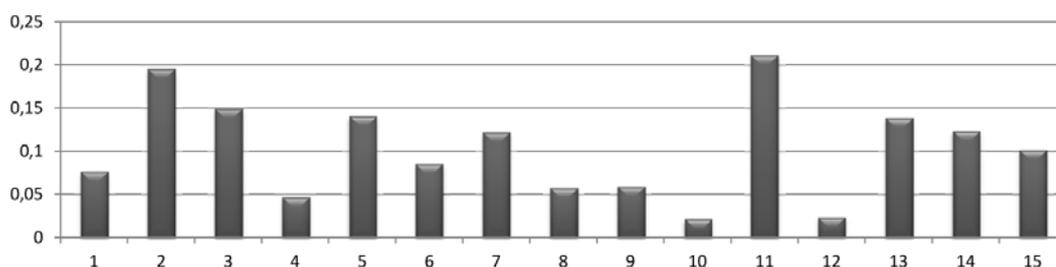


Figure 7 : Densité référentielle de chacun des paragraphes du récit.

Les paragraphes 4, 10 et 12 présentent une densité relativement faible, ce qui s'explique par les référents mentionnés : l'immeuble dans le §4, le chantier dans le §10 et le canal dans le §12 (la figure 7 reflète bien la densité référentielle des personnages humains uniquement). Dans une moindre mesure, les paragraphes 8 et 9 présentent également une faible densité référentielle, du fait des référents principaux qui sont aussi des lieux et non des personnes. Le paragraphe avec la densité référentielle la plus élevée est le §11, ce qui, là encore, présente une explication immédiate : la présence d'un dialogue, avec des termes d'adresse (« Fabre », « tu simules, Jacqueline », « Monte, Paul ») et des pronoms personnels comptabilisés comme maillons de chaînes de coréférence, par exemple « j' » dans « j'ai du lait ». Pour mieux appréhender les variations de références au long du texte, reprenons les titres suggérés pour chacun des paragraphes par (Fuchs *et al.*, ce volume) et ajoutons, sur la base des densités référentielles de chaque personnage, le ou les personnages les plus mentionnés. Un personnage unique est retenu s'il y a un écart important entre celui-ci et le deuxième personnage le plus mentionné ; plusieurs personnages sont retenus s'ils ont à peu près la même densité référentielle dans le paragraphe concerné. La table 3 ainsi obtenue montre tout d'abord l'importance du cercle familial, au début du récit, puis le remplacement progressif de la mère par son effigie. Dans la deuxième partie du récit, chaque paragraphe met surtout en avant un personnage, soit le père, soit le fils, avant de les regrouper. Ce regroupement se fait en deux temps : coexistence des deux référents individuels (§14), puis référent pluriel (§15), ce qui illustre bien l'importance de ce référent de type groupe.

Par ailleurs, quelques paragraphes parlent de personnages totalement différents de ceux du cercle familial : le §4 de l'artiste Flers, le §7 de l'usager de l'espace vert, et le §11 de Jacqueline. On retrouve les trois personnages secondaires qui forment eux aussi des chaînes de coréférence. Ce qui est remarquable, c'est qu'à chaque fois qu'un tel personnage secondaire apparaît, il devient le référent le plus dense au niveau du paragraphe, et – avec une

légère exception concernant Flers qui est mentionné également au §5 – au niveau d’un paragraphe exclusivement. Autrement dit, chaque paragraphe semble focalisé sur un ou un ensemble réduit de personnages bien précis. Concernant Flers, l’usager ou Jacqueline, c’est ainsi un personnage humain qui permet à l’auteur d’introduire un flashback ou un rebondissement dans l’histoire, tel que le dépérissement de l’immeuble ou les retrouvailles du père et du fils : les référents humains servent de chevilles narratives. Cette focalisation des paragraphes sur un ou quelques personnages bien délimités est visible sur les figures 8 et 9, qui illustrent graphiquement les nombres de formes explicites et les densités référentielles personnage par personnage, paragraphe par paragraphe. Sur les deux figures, on peut ainsi constater que l’usager et Jacqueline (et Flers dans une moindre mesure) ressortent avec une seule barre chacun. L’usager en particulier provoque une véritable coupure dans les deux graphiques : il n’apparaît que dans le §7, et c’est le seul personnage humain du §7.

Paragraphe	Titre (cf. Fuchs <i>et al.</i> , ce volume)	Personnages centraux
§1	Incendie et réinstallation du père et du fils	père, fils, mère
§2	Nouvelle vie du père et du fils à l’intérieur	père, fils, mère
§3	Nouvelle vie du père et du fils à l’extérieur	père, fils, effigie
§4	L’immeuble Wagner et l’image de la mère (flashbacks)	Flers
§5	L’image de la mère (flashbacks), retour au père et au fils	père, fils, mère, effigie
§6	Fin de la vie commune du père et du fils ; démolition	père, fils, effigie
§7	Dépérissement de l’espace vert	usager
§8	Dégradation du lieu et de l’image de la mère	effigie
§9	Construction progressive d’un nouvel immeuble	fils
§10	Fin des visites du fils	fils
§11	Retrouvailles du fils avec le père, installé...	Jacqueline
§12	Dans le nouvel appartement du père	père
§13	Flashback sur l’emménagement du père	père
§14	Retour du fils pour le week-end	père, fils
§15	Déjeuner, puis attaque du grattage	père et fils

Table 3 : Personnages principaux, paragraphe par paragraphe.

Dans les figures 8 et 9 on constate également un entrecroisement des personnages principaux, comme cela est classique dans les textes narratifs. La ligne relative au personnage de la mère ainsi que celle relative à l’effigie présentent une particularité (par rapport aux lignes du père et du fils) : celle de diminuer petit à petit. Ni la mère ni l’effigie ne sont mentionnées dans les deux derniers paragraphes, alors que ce sont bien elles qui justifient les actions du récit (cf. Charolles, ce volume, fin du paragraphe 2.2). Par ailleurs, parmi les barres les plus élevées dans les graphiques se trouve celle du groupe formé par le père et le fils dans le dernier paragraphe : non seulement la transition de deux référents individuels à un référent collectif est flagrante, mais, qui plus est, elle conduit à l’une des densités référentielles les plus élevées du récit. La fréquence importante du pronom « on » est ici en cause.

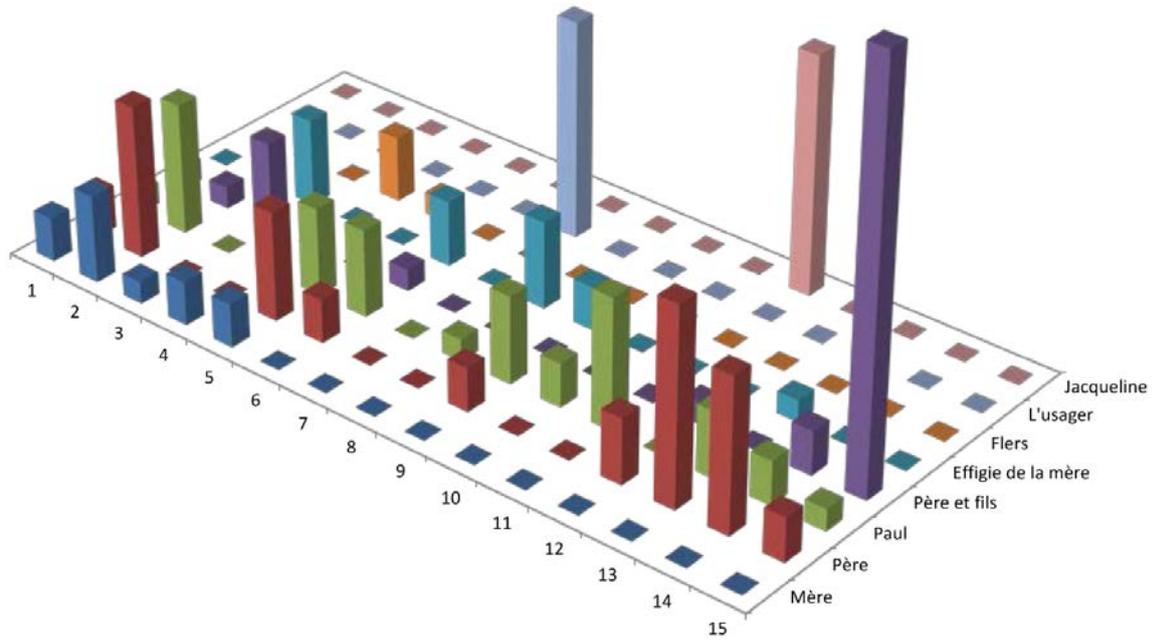


Figure 8 : Nombres de références aux personnages principaux, paragraphe par paragraphe.

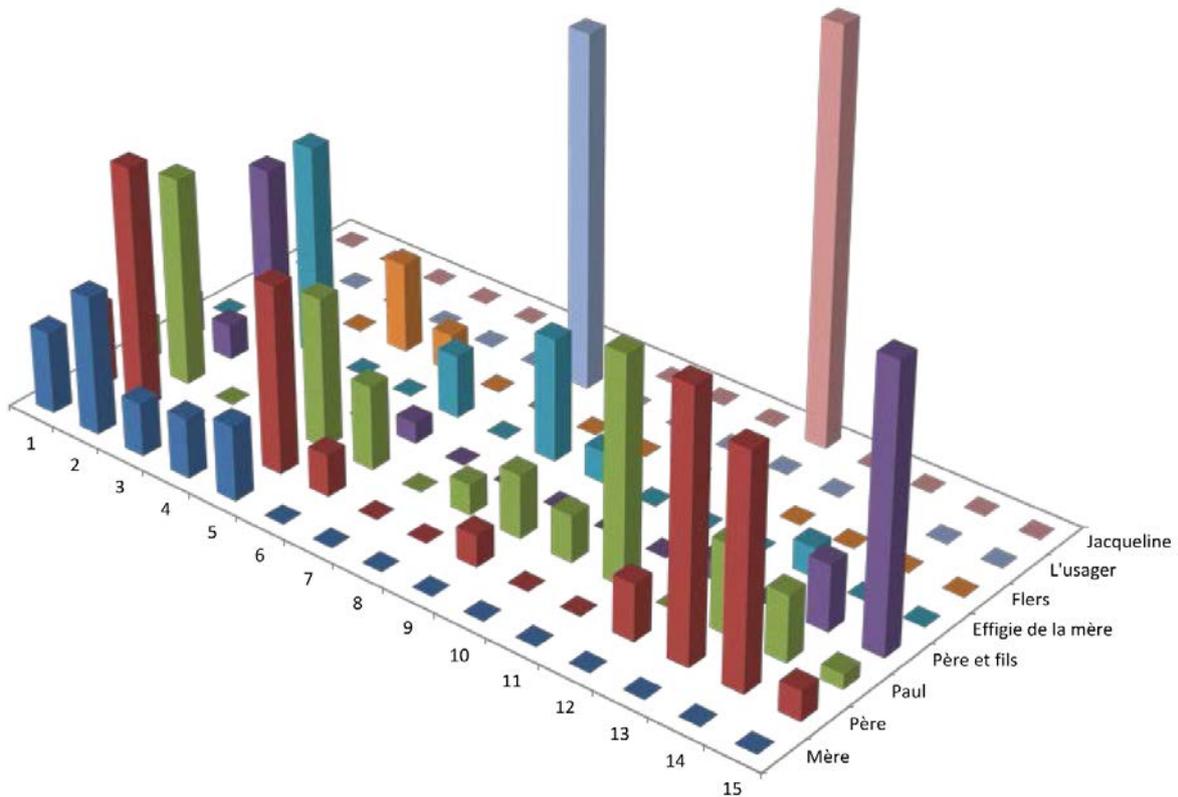


Figure 9 : Densités référentielles, paragraphe par paragraphe.

En se focalisant sur les paragraphes 2 à 5, la figure 10 montre une représentation de la suite des références du texte sous la forme d'une chaîne de points colorés. La figure met en

fait deux représentations en parallèle : la première n'affiche que les formes explicites, et la seconde que les formes atténuées. Dans la représentation des formes explicites, les couleurs ont été choisies pour coder les catégories de formes explicites : bleu foncé pour un groupe nominal défini, bleu clair pour un groupe nominal possessif, rouge pour un déterminant possessif, etc. Dans la partie supérieure de la figure, on peut ainsi appréhender visuellement les alternances de catégories, via les alternances de couleurs. Le chiffre indiqué au-dessus de chaque point coloré est un indice correspondant au référent : 1 pour la mère, 2 pour le père, 3 pour le fils, 4 pour le groupe composé par le père et le fils, 5 pour le portrait de la mère, 6 pour Flers, etc. L'intérêt de ce type de représentation graphique est d'une part de repérer les successions de référents mentionnés, comme cela apparaîtrait avec la succession « 6 1 6 7 6 6 1 » pour le paragraphe 4, par exemple, et d'autre part – si possible en parallèle – de repérer les successions et alternances typiques de couleurs (et donc de catégories d'expressions). Le paragraphe 4 se distingue ainsi non seulement au niveau du référent principal, mais aussi au niveau des catégories de formes de référence : le bleu foncé figurant un groupe nominal défini y apparaît quasiment une fois sur deux, alors qu'il n'apparaît pas ailleurs dans la portion de texte focalisé. Cette observation souligne le rôle du découpage en paragraphe.

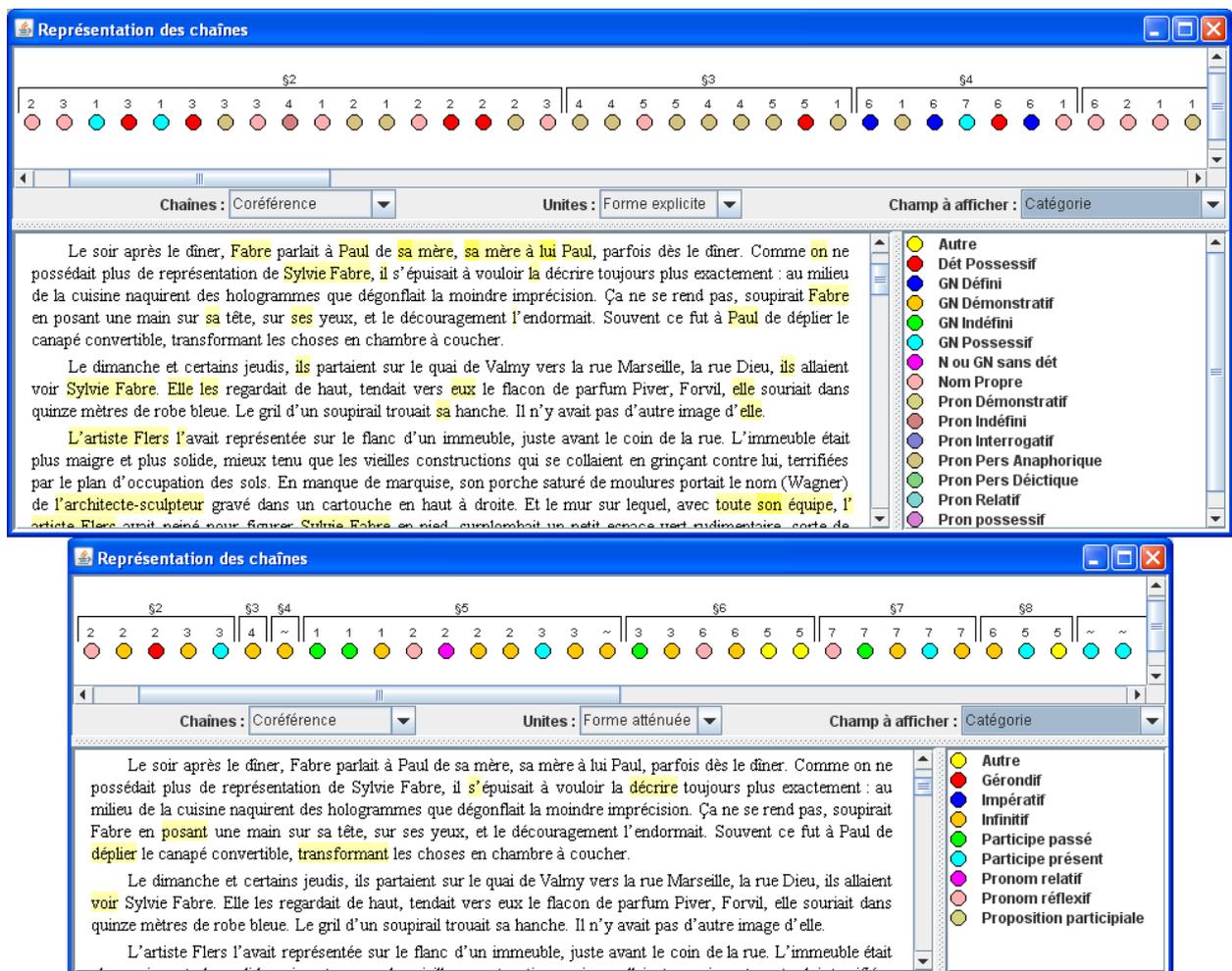


Figure 10 : Une représentation linéaire des suites de références, en haut pour les formes explicites, en bas pour les formes atténuées.

Dans la figure 11, la représentation sous forme de chaîne de points colorés est étendue à l'ensemble du texte. C'est un apport typique de la linguistique outillée : la figure présente les formes explicites et atténuées sous une forme graphique, colorée, dans le but de permettre au linguiste de détecter des tendances, des régularités ou des irrégularités en se reposant sur ses capacités de perception visuelle. Dans cette figure on retrouve les 15 paragraphes du texte, avec les mêmes indices de référents que dans la figure 10 et le même code couleur que dans la représentation des formes explicites de celle-ci. On y observe par exemple que la seule succession de plus de 10 points désignant le même référent n'apparaît qu'à la toute fin du texte, avec le référent 4 qui correspond au groupe formé par le père et le fils. On y observe également qu'un paragraphe ne débute par une référence au dernier référent mentionné dans le paragraphe précédent que deux fois sur l'ensemble du texte, ce qui contredit l'une des hypothèses de (Schneidecker, 1997) sur la structuration textuelle des chaînes de référence (pour un autre genre textuel, il est vrai). Enfin, on y retrouve sous une apparence quelque peu différente les observations réalisées au sujet des densités référentielles.

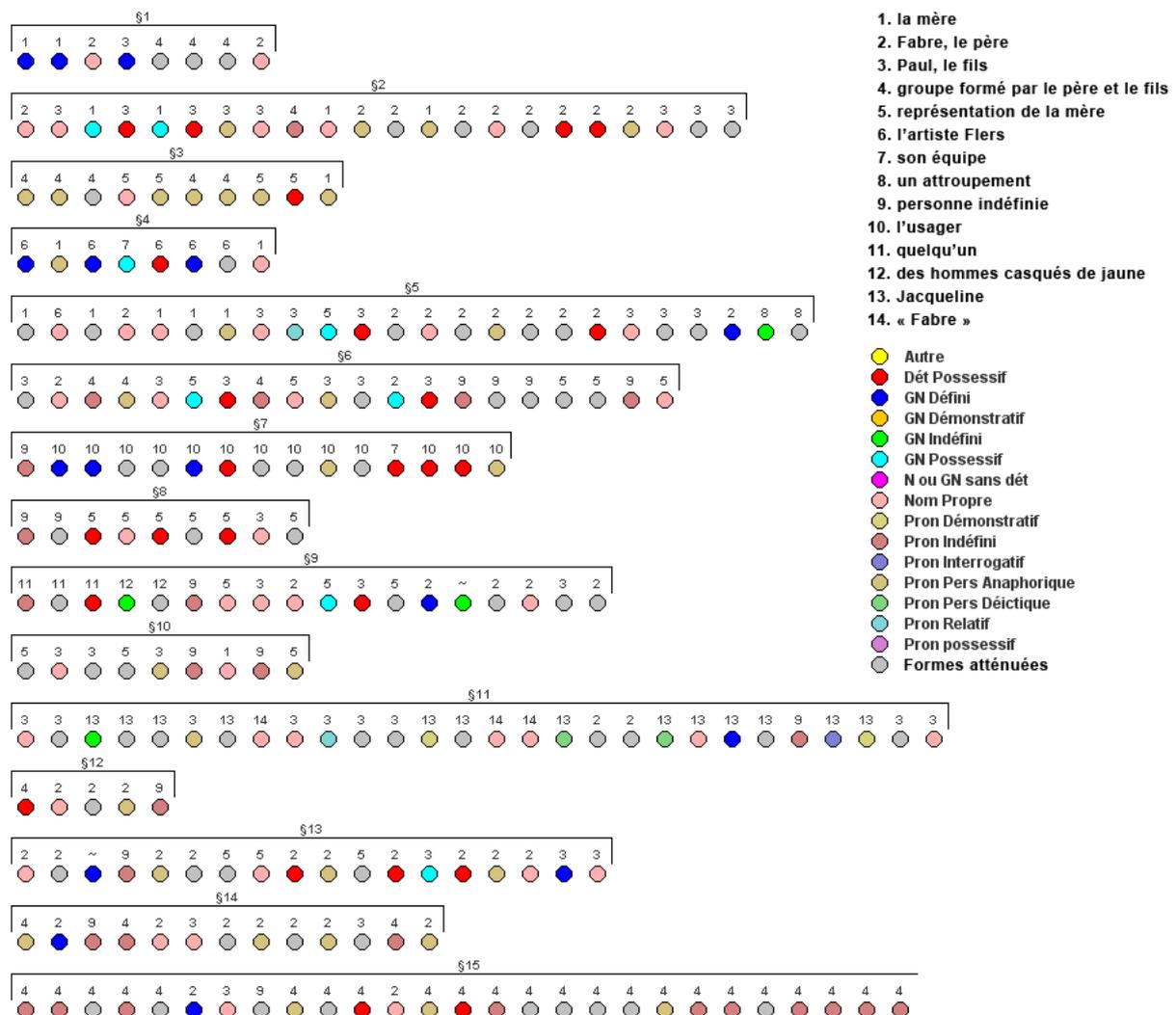


Figure 11 : Représentation graphique de la suite des références, paragraphe par paragraphe, avec visualisation des catégories d'expressions référentielles et prise en compte des formes explicites et des formes atténuées.

6. Conclusion

En complément de l'analyse effectuée dans (Charolles, ce volume), nous avons présenté ici ce que l'on pourrait appeler une « analyse assistée par ordinateur » de la référence aux personnages humains et des chaînes de coréférence afférentes dans le récit de Jean Echenoz. Cette analyse a porté sur trois points principaux. Le premier point a concerné la nature des formes de référence et nous a conduits à distinguer forme explicite et forme atténuée, de manière à questionner la contribution de chacune à l'accession au référent. Il nous a permis de mettre en avant quelques spécificités du récit quant à la nature des référents humains – trois référents principaux clairement définis, des référents secondaires indéfinis ou flous, et la mère avec son effigie au milieu de tout cela – et des expressions référentielles : usage maîtrisé de « on », peu de redénominations, pas de démonstratif et par conséquent peu de proximité psychologique avec les personnages, qui semblent ainsi laisser la primauté aux événements, surtout à la fin du récit. Le second point a concerné la nature des chaînes de coréférence du texte, avec notamment la façon dont elles débutent : de manière classique pour le père et le fils, de manière originale pour la mère, pour laquelle un mélange et une confusion entre humain et non humain sont entretenus tout au long du texte, et ce dès sa toute première phrase. Le troisième point a concerné les rapports entre références et structure du récit, avec quelques observations portant sur la manière dont chaque paragraphe s'articule autour d'un référent ou d'un ensemble de référents précis, avec quelques paragraphes dédiés à un référent qui n'apparaît nulle part ailleurs. Une étude des référents non humains du récit devrait probablement confirmer cette dernière remarque. La fin de l'histoire illustre de manière significative ces transitions parfois brusques d'un référent à l'autre, avec sa transition de deux référents individuels que sont le père et le fils vers un référent collectif, matérialisé par l'usage de « on », qui regroupe père et fils dans une même entité chargée du fameux grattage final.

Références

- Avanzi, Lacheret-Dujour & Victorri (2008). ANALOR, A Tool for Semi-Automatic Annotation of French Prosodic Structure, In: *Proceedings of the 4th Conference on Speech Prosody*, Campinas, Brésil, pp. 119-122.
- Charolles (2002). *La référence et les expressions référentielles en français*, Ophrys, Paris.
- Corblin (1995). *Les formes de reprise dans le discours. Anaphores et chaînes de référence*, Presses Universitaires de Rennes, Rennes.
- Cunningham, Maynard, Bontcheva & Tablan (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics*, p. 168-175.
- Fort (2012). Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus, Thèse de Doctorat, Université de Paris 13.
- Habert (2005). *Instruments et ressources électroniques pour le français*, Ophrys, Paris.
- Habert, Nazarenko & Salem (1997). *Les Linguistiques de corpus*, Armand Colin, Paris.
- Landragin (2011), Une procédure d'analyse et d'annotation des chaînes de coréférence dans des textes écrits, *Corpus 10*, pp. 61-80.
- Landragin, Poibeau & Victorri (2012). ANALEC: a New Tool for the Dynamic Annotation of Textual Data, In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, Istanbul, Turquie, pp. 357-362.
- Mathet & Widlöcher (2013). Glozz Annotation Platform, <http://www.glozz.org/>
- Morton & Lacivita (2003). WordFreak: An Open Tool for Linguistic Annotation. In *Proceedings of Human Language Technology (HLT) and North American Chapter of the Association for Computational Linguistics (NAACL)*, p. 17-18.

- Müller & Strube (2001). MMAX: A tool for the Annotation of Multi-modal Corpora. In: *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Seattle, pp. 45-50.
- Schnedecker (1997). *Nom propre et chaîne de référence*, Klincksieck, Paris.
- Victorri (2013). ANALEC download Web page, <http://www.lattice.cnrs.fr/Analec>
- Widlöcher & Mathet (2009). La plate-forme Glozz : environnement d'annotation et d'exploration de corpus. In: *Actes de la 16e Conférence Traitement Automatique des Langues Naturelles (TALN)*, Senlis.