hal-00279663, version 1 - 15 May 2008

Dominique Labbé

Institut d'Etudes Politiques de Grenoble
BP 45
38040 Grenoble Cedex

dominique.labbe@iep-grenoble.fr

Romain Gary et Emile Ajar

(mai 2008)

De 1945 à 1980, R. Gary a publié une trentaine de romans et essais sous son nom ou sous pseudonyme. Son premier roman (*Education européenne*) a obtenu le prix Interalliés en 1945. Il a également reçu deux fois le prix Goncourt : la première en 1956 pour les *Racines du ciel* (publié sous son nom) et une seconde fois en 1975 pour la *Vie devant soi*, paru sous le pseudonyme de E. Ajar.

Pour l'histoire de cette mystification littéraire organisée par R. Gary avec la complicité de son cousin Paul Pavlowitch, voir : Gary 1981[1], Pavlowitch 1981, Bona 1987.

Michel lafon et Benoît Peeters ont consacré un chapitre de leur livre *Nous est un autre* (2006) à cette mystification. Pour écrire ce chapitre, ils ont souhaité savoir si réellement R. Gary était parvenu à créer un vocabulaire et un style propres à Ajar - distincts des siens – ce qui expliquerait que tous les critiques littéraires ont été trompés. Ou bien si des méthodes scientifiques d'attribution d'auteur auraient permis de démasquer R. Gary.

En réponse à ces questions, le 5 juin 2004, la note ci-dessous a été envoyée à Michel lafon et Benoît Peeters qui l'ont partiellement utilisée dans leur ouvrage.

En mai 2008, les tableaux de cette note ont été modifiés pour intégrer dans les calculs le dernier roman de R. Gary (les *Cerfs-volants*, 1980) qui n'avait pas été pris en compte en 2004.

[1] Les références complètes des ouvrages cités sont données en annexe.

hal-00279663, version 1 - 15 May 2008

Grenoble 5 juin 2004

Vous avez bien voulu m'interroger sur le cas Gary-Ajar qui semble intéresser encore le public comme en témoigne la réédition récente de la biographie de cet auteur (Bona 1987).

Est-ce que la statistique permet d'identifier Gary dans l'ombre de Ajar, avec quels outils et avec quel degré de certitude ?

Comme indiqué dans mes précédents courriers, et comme l'explique l'essai sur *Corneille dans l'ombre de Molière*, la principale difficulté réside dans l'impossibilité d'accéder aux fichiers électroniques de la quasi-totalité des romans français du XXe siècle. Certes, ces fichiers existent à l'Institut de la Langue Française, mais cet organisme refuse de communiquer les œuvres couvertes par des copyrights, même aux fins de recherche. Il fallait donc passer ces livres au scanner, ce qui est un travail très long du moins si l'on veut un traitement sans faute…

Pour cette raison, la note ci-dessous se limite à une comparaison des 4 romans signés Ajar avec les cinq romans contemporains de Gary. Pour Ajar, les 4 romans ont tous été publiés au Mercure de France : *Gros-Câlin* (1974), *La vie devant soi* (1975), *Pseudo* (1976) et *L'angoisse du roi Salomon* (1979). Les 5 romans contemporains signés Gary, ont tous paru chez Gallimard : *Chien blanc* (1970), *Au-delà de cette limite, votre ticket n'est plus valable* (1974), *Charge d'âme* (1977), *Clair de femme* (1977), les *Cerfs-volants* (1980). Au total ce corpus compte 510 180 mots, soit 200.000 mots pour Ajar et 300.000 mots pour Gary.

Après passage au scanner, les textes ont été normalisés et lemmatisés, selon des procédés éprouvés sur lesquels on ne revient pas ici (Labbé, 1990), avant de subir les traitements statistiques standards.

Parmi les traitements intéressants pour l'attribution d'auteur, cette note présente les résultats du calcul de la distance intertextuelle et l'observation de certaines combinaisons caractéristiques de mots (verbe + verbe).

I. La distance intertextuelle

L'ensemble du raisonnement est présenté dans les articles parus en décembre 2001 dans le *Journal of Quantitative Linguistics* et, en français, dans le numéro 2 de la revue *Corpus* (2003). Il s'agit de mesurer la proximité ou l'éloignement de deux ou plusieurs textes, les uns par rapport aux autres, en **considérant tous leurs mots**. On calcule ainsi un indice de la distance qui varie uniformément entre zéro (exactement les mêmes mots avec les mêmes fréquences) et un (aucun mot en commun).

Quatre facteurs influencent cette distance :

— le **genre**. On ne parle pas comme on écrit ; la fiction romanesque a ses codes qui ne sont pas ceux du théâtre, etc. Le carcan imposé par le genre est plus ou moins rigide. Celui de la fiction romanesque à l'époque contemporaine étant très lâche, la distance aura tendance à être plus forte que dans d'autres genres plus codifiés ;

— le vocabulaire de **l'époque**. L'œuvre de Gary s'étend sur plus de 35 ans. Sur de telles étendues de temps, le style et le vocabulaire d'un auteur évoluent nécessairement. D'ailleurs, il est lui-même pris dans le flux qui change lentement le lexique de la langue. Il est donc indispensable de comparer des œuvres contemporaines, ce qui est fait ici.

— le **thème** traité. Chaque thème a un vocabulaire propre. Ce sont d'abord des noms de lieux, de personnes et une série de substantifs particuliers…

— enfin et surtout : **l'auteur**…

hal-00279663, version 1 - 15 May 2008

Pour rechercher la paternité d'un texte anonyme ou dont l'auteur est contesté, il faut donc le comparer à d'autres dont la signature n'est pas contestée, ayant été écrits à la même époque et traitant de thèmes voisins, dans un même genre (poésie, prose, roman, théâtre…). Ce point est important. En l'état actuel de la technique, on ne peut comparer que ce qui est comparable : un roman avec un autre roman, le théâtre avec le théâtre, une lettre avec de la correspondance…

Ce calcul - appliqué à plusieurs milliers de textes en français de toutes origines (romans, pièces de théâtre, poésie, articles de presse, discours politiques, entretiens…) depuis le XVIIe siècle - a permis de confirmer la validité du raisonnement et d'étalonner une échelle de la distance. Les valeurs ci-dessous s'appliquent aux textes dont les longueurs sont comprises entre 5 000 et 20 000 mots pour les textes plus longs, on utilise des extraits selon la méthode présentée dans Labbé (2007) :

— une valeur inférieure ou égale à 0.20 ne se rencontre jamais chez des auteurs différents ;

— entre 0.20 et 0.25, il est pratiquement certain que l'auteur est le même. Sinon, les deux textes ont été écrits à la même époque, dans un même genre, sur un sujet et avec des arguments semblables. Ce cas se rencontre parfois dans les articles de presse, à propos d'un même événement, parce que les journalistes travaillent à partir des mêmes sources et citent les mêmes noms de lieux et de personnes... Dans le cas d'œuvres littéraires appartenant à deux auteurs différents, il est très probable que le second s'est plus qu'"inspiré" du premier (dans l'ordre chronologique). En tous cas, ce genre de "collision" peut difficilement se produire plusieurs fois entre deux auteurs distincts.

— au-dessus 0.25, on entre dans une zone "grise" où deux hypothèses sont envisageables : un même auteur et des thèmes très différents ou deux auteurs contemporains traitant, dans un même genre, un thème proche avec leur style propre… De telle sorte que, plus on s'élève au-dessus de ce seuil, plus il sera difficile d'attribuer la paternité d'un texte anonyme à l'auteur considéré sans que, pour autant, cette paternité puisse être rejetée ;

— au-dessus de 0.40 les auteurs sont très probablement différents ou bien, pour un même auteur, les textes sont de genres très éloignés, par exemple : oral et écrit.

Les trois tableaux ci-dessous présentent les résultats détaillés obtenus sur l'ensemble du corpus (les résultats significatifs sont placés en gras). Les deux premiers portent sur les œuvres signées Ajar puis Gary qui sont comparées dans le troisième.

Tableau I. Les distances entre les 4 romans de Emile Ajar

	Gros-Câlin	Vie devant soi	Pseudo	Salomon	Moyenne
Gros câlin		0,273	**0,224**	**0,247**	**0,248**
Vie devant soi	0,273		0,255	**0,176**	**0,235**
Pseudo	0,224	0,255		**0,246**	**0,242**
SalomonTotal	0,247	0,176	0,246		**0,223**

* Ce tableau est symétrique selon la diagonale (comme les deux suivants). On ne considère que la moitié nord-est.

D'après l'échelle des distances présentée ci-dessus, on en tire que les 4 romans ont bien été écrit par la même personne et que deux d'entre eux sont même très proches (la *Vie devant soi* et *l'Angoisse du roi Salomon*). En fait, cette distance très faible concerne le dernier Ajar paru et montre que Gary, du moins après *La vie devant soi*, avait "codifié" le style et le vocabulaire propre à "Ajar" et qu'il était capable de le reproduire assez fidèlement…

La même conclusion peut être tirée des distances entre les cinq romans publiés à la même époque sous le nom de R. Gary (tableau II) mais la dispersion est plus grande et, surtout, la dernière colonne du tableau indique que deux romans sont un peu "décalés" par rapport aux trois autres (*Charge d'âme* et *Clair de femme*). Enfin, il n'est pas indifférent de constater que

le dernier roman – les Cerfs-volants publié peu avant son suicide – est aussi le plus central de son œuvre officielle.

Tableau II. Les distances entre les 5 romans de Romain Gary contemporains de ceux de Emile Ajar.

	Chien blanc	Au-delà…	Charge…	Clair de f…	Cerfs-v…	Moyenne
Chien blanc		**0,227**	0,263	0,267	**0,224**	**0,245**
Au-delà…	0,227		0,280	**0,229**	**0,205**	**0,235**
Charge…	0,263	0,280		0,314	**0,249**	0,276
Clair de f…	0,267	0,229	0,314		**0,240**	0,263
Cerfs-v…	0,224	0,205	0,249	0,240		**0,230**

Enfin, le tableau III compare les 5 romans de R. Gary avec les 4 parus sous le pseudonyme de E. Ajar et met en valeur un certain nombre d'anomalies, du moins dans l'optique de deux auteurs différents (en gras dans le tableau)

Tableau III. Les distances entre les 5 romans de Romain Gary et les 4 romans parus sous le nom de Emile Ajar.

	Chien blanc	Au-delà…	Charge…	Clair de f…	Cerfs-v…	Moyenne
Gros câlin	0,255	**0,238**	0,311	**0,244**	0,252	0,260
Vie devant soi	0,327	0,310	0,383	0,275	0,309	0,321
Pseudo	0,266	**0,247**	0,323	**0,238**	0,256	0,266
SalomonTotal	0,305	0,263	0,360	**0,226**	0,270	0,284
Moyenne	0,288	0,264	0,344	**0,246**	0,272	0,283

Avant d'examiner les "anomalies" contenues dans le tableau III, une expérience de classification automatique[2] permet de comprendre le problème posé par ces données (tableau IV).

Cette classification "arborée" vise à représenter aussi fidèlement que possible les proximités et les éloignements relatifs de chaque texte par rapport à tous les autres. La distance entre deux textes est symbolisée par la longueur des chemins à parcourir pour relier les deux "feuilles" correspondantes sur l'arbre. Par exemple, on voit que les deux textes les plus éloignés sont : *La vie devant soi* (Ajar, 1975) et *Charge d'âmes* (Gary, 1977) et que les deux plus proches sont *La vie devant soi* et *l'Angoisse du roi Salomon* (Ajar).

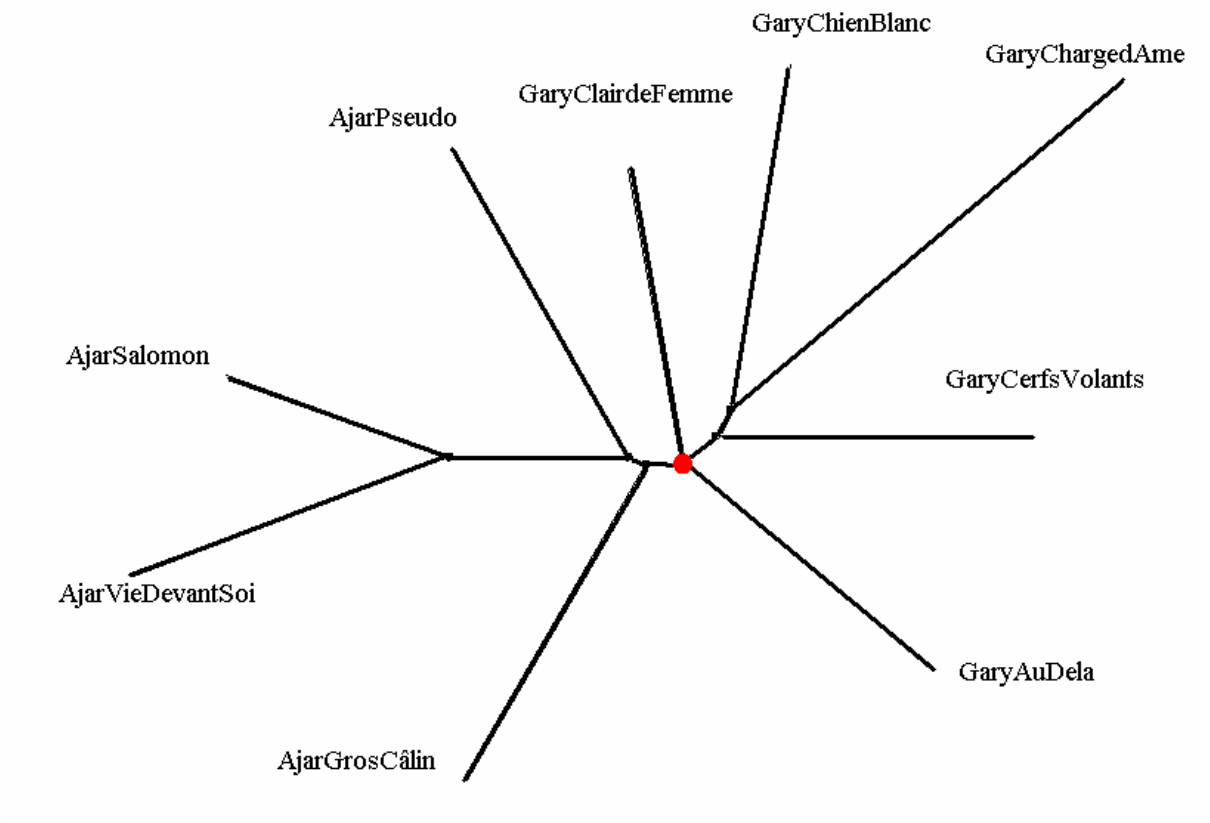
Un examen superficiel du graphique peut conduire à conclure à deux auteurs différents : Ajar à gauche et Gary à droite. En fait, le diagramme indique également que deux romans de Gary (*Clair de femme* et *Au-delà de cette limite*…) sont plus proches de deux romans de Ajar (*Pseudo* et *Gros câlin*) que des autres textes parus sous le nom de Gary (ce qu'indique déjà la comparaison des valeurs correspondant à ces deux textes dans les tableaux II et III).

[2] Pour une présentation de cette classification arborée : Luong 1988 et Ruhlmann 2003. L'algorithme de classification utilisé pour tracer l'arbre du tableau IV est présenté dans ce second ouvrage qui est consultable en ligne.

hal-00279663, version 1 - 15 May 2008

hal-00279663, version 1 - 15 May 2008

Tableau IV. Classification arborée sur le corpus "Gary-Ajar"



Pour rendre les résultats plus intelligibles, les plus faibles distances sont classées par ordre croissant (tableau IV).

Tableau IV. Les couples les plus proches
(en gras les distances entre des œuvres signées sous deux noms différents)

1	Ajar Vie devant soi	Ajar Salomon	0,176
2	Gary Au-delà de cette limite…	Gary Cerfs-volants	0,205
3	Ajar Gros câlin	Ajar Pseudo	0,224
4	Gary Chien blanc	Gary Cerfs-volants	0,224
5	**Ajar Salomon**	**Gary Clair de femme**	**0,226**
6	Gary Chien blanc	Gary Au-delà de cette limite…	0,227
7	Gary Au-delà de cette limite…	Gary Clair de femme	0,229
8	**Ajar Gros câlin**	**Gary Au-delà de cette limite…**	**0,238**
9	**Ajar Pseudo**	**Gary Clair de femme**	**0,238**

Deux romans de Gary (*Clair de femme* et *Au-delà de cette limite…*), très proches entre eux, sont également très proches des livres de Ajar (*Gros-Câlin*, *Pseudo* et *l'Angoisse du roi Salomon*). Un tel classement croisé entre quatre romans publiés à quelques mois d'intervalle avec des distances aussi faibles ne laisse aucun doute. Ce n'est pas un "accident" ponctuel qui serait nécessairement limité à deux ouvrages. L'auteur est donc le même…

Il faut également considérer l'autre bout de la distribution. N'existe-t-il pas entre ces œuvres des distances telles qu'elles pourraient faire douter de l'existence d'un auteur unique ? Le tableau V apporte la réponse.

Tableau III. Les couples les plus distants
(en gras les distances les plus grandes entre des œuvres signées du même nom)

28	Ajar Salomon	Gary Chien blanc	0,305
29	Ajar Vie devant soi	Gary Cerfs-volants	0,309
30	Ajar Vie devant soi	Gary Au-delà de cette limite	0,310
31	Ajar Gros câlin	Gary Charge d'âme	0,311
32	**Gary Charge d'âme**	**Gary Clair de femme**	0,314
33	Ajar Pseudo	Gary Charge d'âme	0,323
34	Ajar Vie devant soi	Gary Chien blanc	0,327
35	Ajar Salomon	Gary Charge d'âme	0,360

Sans atteindre le seuil de 0,40 au-dessus duquel l'hypothèse de deux auteurs différents devient très probable (pour des textes contemporains appartenant à un même genre), on note les distances très élevées séparant deux Ajar (*La vie devant soi* et *L'angoisse du roi Salomon*) d'un roman signé Gary dont l'écriture est pourtant contemporaine (*Charge d'âme*) mais on note que ce même *Charge d'âme* est aussi éloigné de *Clair de femme* lui aussi contemporain et alors que tous deux sont signés Gary…

Généralisons le raisonnement. Voici ci-dessous les moyennes des distances séparant les 4 romans parus sous son nom (Gary-Gary), celles entre les 4 publiés sous le nom de Ajar (Ajar-Ajar) et enfin celles séparant ces deux ensembles (Gary-Ajar).

Gary-Gary : 0,250
Ajar-Ajar :0,237
Gary-Ajar : 0,285

Sans pouvoir parler, à propos de Gary, d'un "Caméléon" comme le fait l'une de ses récentes biographes, cette dernière mesure souligne malgré tout, une forte diversité dans l'écriture. En fait, dans les dix dernières années de sa vie, il y a bien deux Gary mais la coupure ne passe pas seulement entre Gary et Ajar mais au sein même de Gary.

La coupure au sein de l'œuvre "officielle" séparerait :

— un auteur de facture relativement "classique" dont les œuvres sont écrites à la troisième personne et qui comportent une intrigue, une progression dramatique, des rebondissements… Dans l'ensemble de l'œuvre, ce premier ensemble domine nettement. Dans le corpus, seul *Charge d'âme* se rattache vraiment à ce premier ensemble auquel on peut aussi rattacher la dernière oeuvre : *Les cerfs-volants* (1980).

— une écriture relativement "nouvelle" où la narration est faite à la première personne, dans un ton plus personnel, avec une forme imitant l'oralité. L'auteur y cultive le saugrenu, l'humour noir, le non-sens, le coq-à-l'âne… *La danse de Gengis Cohn* (1966) inaugure probablement cette veine particulière. Dans le corpus, *Chien blanc* (1970) représente ce second versant avant "l'invention de Ajar". Au cours de sa période "ajarienne", Gary signe sous son nom deux romans qui appartiennent manifestement à ce second "genre" : *Au-delà de cette limite…* (1974) et surtout : *Clair de femme* (1977).

La faiblesse des distances séparant ces deux romans de Gary avec les œuvres signées "Ajar" ne laisse aucun doute sur le fait que l'auteur est le même. Surtout cette proximité souligne que Ajar n'est pas une pure création "ex-nihilo". Il existait potentiellement dans ce second versant de l'œuvre de Gary.

Pour conclure définitivement dans ce sens — dualité de l'œuvre de Gary et origine de "Ajar" dans l'une des deux parties de celle-ci —, il faudrait naturellement traiter l'ensemble de l'œuvre de Gary — quel que soit le nom sous lequel les textes ont paru depuis 1945.

hal-00279663, version 1 - 15 May 2008

Si le calcul montre la diversité de l'œuvre signée Gary, il souligne également l'homogénéité de l'œuvre "Ajar". Autrement dit, Gary avait "codifié" l'écriture particulière de celui-ci et il a su s'y tenir dans les quatre livres écrits sous ce nom au cours des cinq années qu'a duré la supercherie. Il est également probable qu'il a essayé de "singulariser" autant que possible cette écriture de sa manière propre d'écrire. Mais il n'y est pas vraiment parvenu comme le suggère d'ailleurs clairement la classification arborée… Cela confirme que *le Ticket* et surtout *Clair de femme* — qui ont été écrits alors que Gary produisait aussi ses Ajar — ont été fortement "contaminés" par ce style si particulier et que l'auteur n'est pas parvenu à maintenir la "cloison" entre ses deux œuvres.

Il est amusant de constater que ce n'est pas Gary qui "contamine" Ajar mais l'inverse puisque *Clair de femme* est postérieur à trois des Ajar (*Gros-Câlin, La vie devant soi* et *Pseudo)* dont il est fort proche…

L'auteur a imité son double ! Je crois me souvenir que D. Bonna signale, dans sa biographie de Gary, que celui-ci aurait été conscient de cet étrange phénomène.

Cette "attribution d'auteur" est-elle confirmée par d'autres indices lexicaux ou stylistiques ?

II. Les syntagmes répétés

Le calcul qui vient d'être présenté considère l'ensemble du vocabulaire. D'autres indices prennent en compte certains aspects plus particuliers. Du point de vue des questions d'attribution d'auteur, l'analyse des "syntagmes répétés" les plus fréquents représente l'une des voies les plus prometteuses.

Ce sont des combinaisons stables de plusieurs vocables qui conservent leur indépendance contrairement au mot composé et à la locution où les composants perdent cette indépendance. Puisqu'on dispose des catégories grammaticales, grâce à la lemmatisation, on peut même s'intéresser à certaines combinaisons particulières comme les combinaisons "verbe +verbe".

Dans la langue française, les "pseudo-auxiliaires", dits encore "verbes modaux", suivis d'un infinitif permettent de réaliser une infinité de groupes verbaux complexes (sur le modèle de "pouvoir faire"). Chaque individu trahit, dans les combinaisons qu'il privilégie, un certain rapport à soi-même, aux autres et aux choses. Le sujet peut envisager ces rapports sous l'angle de l'action ("faire", "aller", "dire"), du possible ("pouvoir"), de la volonté ("vouloir") de l'obligation morale ou légale ("devoir") de l'impératif ("falloir"), de la connaissance ("savoir"). Dans chacune de ces catégories, il existe de multiples nuances (souhaiter, prétendre, entendre, vouloir, exiger…) Ces pseudo-auxiliaires sont suivis d'un verbe à l'infinitif : verbe d'action (faire, aller ou dire), de la pensée (penser, connaître), d'état (être) ou de la possession (avoir).

On demande donc à l'ordinateur de retrouver toutes ces combinaisons en tenant compte de ce que des adverbes ou des locutions adverbiales peuvent se glisser dans le couple recherché (par exemple : "faire (très bien) voir", etc).

Cette méthode a été élaborée avec A. Pibarot en 1997. Elle a été appliquée à plusieurs milliers de textes différents et s'est montrée apte à discriminer les thèmes (grâce aux groupes nominaux) et les auteurs (grâce aux groupes verbaux). Le tableau IV ci-dessous présente les résultats obtenus sur Gary et Ajar.

hal-00279663, version 1 - 15 May 2008

Tableau IV. Les principaux groupes verbaux complexes chez Ajar et chez Gary
(Classement en fonction de la fréquence exprimée pour 10.000 mots)

Ajar Syntagmes	**‰***	**Gary Syntagmes**	**‰***
vouloir dire	**7,20**	**vouloir dire**	**5,80**
pouvoir être	**4,90**	**pouvoir être**	**3,62**
pouvoir faire	**4,40**	**pouvoir faire**	**2,97**
devoir être	**3,10**	**devoir être**	**2,39**
aller faire	**2,90**	**aller faire**	**2,20**
pouvoir vivre	2,70	**devoir avoir**	**2,13**
laisser tomber	**2,30**	**pouvoir dire**	**2,13**
devoir avoir	**2,10**	**laisser tomber**	**1,87**
aller voir	**2,00**	laisser aller	1,55
devoir faire	**1,80**	faire passer	1,49
pouvoir avoir	1,80	**devoir faire**	**1,29**
aller chercher	1,70	entendre parler	1,29
falloir faire	**1,60**	**falloir faire**	**1,29**
falloir être	**1,60**	mettre rire	1,29
vouloir savoir	**1,60**	**aller voir**	**1,23**
aller mourir	1,50	**falloir être**	**1,16**
faire chier	1,50	vouloir savoir	1,16
faire parler	1,50	aller prendre	1,03
falloir savoir	1,50	vouloir être	1,03
pouvoir dire	**1,40**	faire tuer	0,97

* pour 10.000 mots

En ne considérant que les 20 premières combinaisons, Gary et Ajar en partagent 13, notamment les cinq premières dans le même ordre hiérarchique et pratiquement avec les mêmes fréquences relatives. Etant donné l'extraordinaire diversité des combinaisons possibles, il est impossible de rencontrer deux individus présentant les mêmes préférences pour plusieurs combinaisons spécifiques, à peu près dans le même ordre et avec des densités voisines. Certes, on pourrait objecter que, par exemple, "vouloir", "pouvoir", "dire", "être" et "faire" sont des verbes très fréquents et que leurs combinaisons sont assez banales — en fait, même à ce niveau-là, les préférences individuelles se distinguent très bien — mais alors que dire de "laisser tomber" par exemple ? Cette combinaison se trouve en septième position chez Ajar, en huitième chez Gary avec des densités comparable.

L'étude des syntagmes répétés chez R. Gary montre que ce qui caractérise cet auteur, sous son nom comme sous celui de Ajar, c'est l'insistance sur le "dire" et sur "parler", dimension qui se retrouve par exemple dans l'importance que les dialogues prennent dans l'œuvre de Gary dès l'origine, importance que l'on retrouve également chez Ajar...

Naturellement, certaines de ces combinaisons appartiennent au thème traité dans l'un ou l'autre des romans. Par exemple, le vingtième syntagme chez Gary ("faire tuer") appartient en quasi-totalité à *Charge d'âme*. D'autres, comme "faire chier" ont un poids très marginal chez Gary : ce n'est qu'à l'abri du nom d'un autre, qu'il peut donner libre cours à un vocabulaire "scatologique" dont il limite assez strictement l'emploi dans son œuvre "officielle"…

En conclusion.

Les quatre romans signés Ajar sont en décalage assez net par rapport à l'œuvre "classique" de Gary, celle pour laquelle il était surtout connu lorsque Ajar paraît sur la scène littéraire. Cela peut expliquer que l'on n'ait pas songé à lui avant que son lien de parenté avec

hal-00279663, version 1 - 15 May 2008

Pavlowitch n'ait été révélé. Si Gary n'avait pas donné *Chien blanc* et surtout *: Au-delà de cette limite*… et *Clair de femme*, le calcul statistique aurait difficilement pu identifier la "plume de l'ombre" cachée derrière le "pseudo" de Ajar. A condition aussi que Gary puis Pavlowitch n'aient pas dévoilé la supercherie, il aurait été très difficile d'écarter l'idée selon laquelle ce dernier était bien l'auteur des œuvres qu'on lui prêtait. Mais, après la parution de *Clair de femme* (1977), le doute n'est plus permis. Les distances intertextuelles désignent un auteur unique : Gary…

A la fin des années 1970, les ordinateurs disposaient déjà d'une puissance respectable et les outils intellectuels nécessaires pour "démasquer" Ajar, existaient potentiellement. Pourtant, personne, à notre connaissance, n'a songé à utiliser la statistique lexicale pour résoudre cette énigme, y compris parmi les chercheurs. L'explication essentielle réside dans la croyance, encore largement partagée aujourd'hui, selon laquelle "l'auteur est mort". En fait, non seulement l'auteur est bien présent dans les œuvres qu'il publie sous son nom, mais de plus — même lorsqu'il s'agit d'un vieux routier comme Gary — il lui est très difficile de dissimuler longtemps les traces qu'il laisse malgré lui dans les textes qu'il écrit pour d'autres (réels ou imaginaires !)

Laissons le dernier mot à R. Gary :

"Je ne crois pas qu'un «dédoublement» soit possible. Trop profondes sont les racines des œuvres, et leurs ramifications, lorsqu'elles paraissent variées, très différentes les unes des autres, ne sauraient résister à un véritable examen et à ce qu'on appelait autrefois «l'analyse des textes»" (Gary, 1981, p. 34)

Références bibliographiques

Bona Dominique (1987), *Romain Gary*, Paris, Mercure de France.

Gary Romain (1981), *Vie et mort de Emile Ajar*, Paris, Gallimard.

Labbé Dominique, (1990), *Normes de saisie et de dépouillement des textes politiques*, Grenoble, Cahier du CERAT.

Labbé Cyril, Labbé Dominique (2001). "Inter-Textual Distance and Authorship Attribution Corneille and Molière". *Journal of Quantitative Linguistics*. 8-3. December 2001.

Labbé Cyril, Labbé Dominique (2003). "La distance intertextuelle". *Corpus*. 2-2003.

Labbé Dominique (2003). *Corneille dans l'ombre de Molière. Histoire d'une recherche*. Bruxelles, Les impressions nouvelles.

Lafon Michel & Peeters Benoît (2006). *Nous est un autre*. Paris, Flammarion.

Luong Xuan (1988). *Méthodes d'analyse arborée. Algorithmes, applications.* Thèse pour le doctorat ès sciences. Paris, Université de Paris V.

Pavlowitch Paul (1981), *L'homme que l'on croyait*, Paris, Fayard.

Pibarot André, Labbé Dominique (1998) "Les syntagmes répétés dans l'analyse des commentaires libres", in Sylvie Mellet (dir), *4e Journées d'analyse des données textuelles*, Nice, p 507-516.

Ruhlmann Mathieu (2003), *Analyse arborée. Représentation arborée par la méthode des groupements*, Rapport de stage sous la direction de LABBE (Cyril) et LABBE (Dominique), Grenoble : Polytech'Grenoble et Cerat-IEP, août 2003.

hal-00279663, version 1 - 15 May 2008