



Villa R

Paul Klee, 1919.

Profils linguistiques et structure textuelle

Mathilde Dagnat
LPL, Aix-en-Provence

Plan

- 1 Problématique générale et hypothèses
- 2 Traitement des données
- 3 Interprétation stylistique des résultats
- 4 Conclusions et prolongements



1. Problématique générale et hypothèses

1.1. Étude générale

Représentation de l'oralité populaire québécoise (OPQ) dans cinq pièces de théâtre de Michel Tremblay (sous-corpus *tremblay*) et son évolution entre 1968 et 1998 (Dargnat 2006).

1.2. Deux hypothèses principales

1. *Le réalisme.* Les saillances linguistiques dans les textes de théâtre renvoient à des observables linguistiques caractéristiques de l'OPQ.

2. *L'évolution.* La fonction et la distribution de ces marques dans les pièces ont évolué entre 1968 et 1998.

1.3. Trois étapes

1. Définition de l'objet OPQ

Construction du référent = liste de traits dans la littérature et corpus linguistique de référence *frcapop*, à partir des corpus *Sankoff-Cedergren* et *Montréal 84* (cf. Thibault & Vincent 1990).

2. Analyse comparative du corpus de référence et du corpus littéraire

Corpus *frcapop* et corpus *tremblay*.

Perspectives phonographique, sémantico-pragmatique, syntaxique et lexicale.

3. Interprétation stylistique

Correspondance entre identité narrative et profil linguistique des personnages et évolution de cette correspondance (par rapport aux modèles esthétiques québécois).

1.4. Analyse proposée ici

La première hypothèse est considérée comme acquise.

C'est l'hypothèse de l'évolution dans *tremblay* qui va être validée et décrite, notamment au moyen de la catégorisation des profils linguistiques + / – OPQ.

Rem. Je ne traite donc pas ici du corpus *frcapop*, mais de **deux pièces** du corpus *tremblay*, qui permettent d'asseoir l'analyse.

Les Belles-sœurs (1968)

Bonjour, là, bonjour (1974)

***L'Impromptu d'Outremont* (1980)**

Le Vrai Monde ? (1987)

***Encore une fois, si vous permettez* (1998)**

Les données et les résumés des cinq pièces figurent dans l'exemplier (p.3-12).



2. Traitement statistique des données

2.1. Exploitation logicielle du corpus *tremblay*

- Étape 1* Numérisation des cinq textes (scannage et relecture).
- Étape 2* Codage XML en suivant (en partie) les consignes de la la TEI.
Corpus non taggé pour des raisons de temps de faisabilité
automatique (formes non standard et non systématiques).
- Étape 3* Exploitation avec le logiciel *Weblex* (Heiden 2002).
<http://weblex.ens-lsh.fr/wlx>

Avantages

- Systématicité et répétabilité des requêtes (CQP).
- Sélection du matériau verbal des locuteurs-cibles (L2 et personnages).

Inconvénients

- Pas de récupération des phénomènes non marqués morphologiquement (dislocations, parataxes, accords sémantiques, etc.) (Dargnat 2006, 2008).
- Pas de fonction pour effectuer des calculs statistiques permettant de tester des hypothèses de significativité (par ex. Khi 2).

2.2. Variables du profil linguistique + / – OPQ

[Exemplier A, p. 2]

Néographies phonétisantes (Anis : 1999)

- a. *pis*
- b. *ben*
- c. *toé* et *moé*
- d. Ouverture de [ɛ] en [a] devant [r] (ex. *farmer*, *narveuse*, etc.)
- e. Mots terminés par ou comportant une apostrophe devant consonne (ex. *v'nir*, *a'voulait*, etc.)

Phénomènes syntaxiques

- f. Prép. *dans* et *à* devant nom sans déterminant (ex. *dans cuisine*, *à matin*)
- g. Particule *ça fait que*
- h. Particule interrogative et exclamative *tu* (ex. *ils viennent-tu ?*)
- i. Négations en *pas* sans *ne* (ex. *j'en veux pas*)

Phénomènes lexicaux

- j. Sacres (ex. *câlisse, maudit, tabarname*, etc.)
- k. Particularités dialectales ou « québécoisismes » (ex. *fifi, siau, écrapoutir, garrocher*)
- l. Anglicismes (ex. *turtle neck, chambre de bain, braidage, smatte, toune, bécosses*)

Le choix des variables peut être discuté. Il résulte de multiples contraintes.

Avantage : une couverture multidimensionnelle.

Inconvénient : un comptage semi-automatique.

Je reviens sur ce point en conclusion.

2.3. Valeurs numériques des variables

L'Impromptu d'Outremont (1980)

Données brutes

IO	<i>a.</i>	<i>b.</i>	<i>c.</i>	<i>d.</i>	<i>e.</i>	<i>f.</i>	<i>g.</i>	<i>h.</i>	<i>i.</i>	<i>j.</i>	<i>k.</i>	<i>l.</i>
Fernande	1	0	0	0	16	0	0	0	24	0	13	14
Lorraine	23	8	0	0	46	3	0	4	40	0	12	9
Lucille	2	0	0	0	45	0	0	0	44	0	8	16
Yvette	1	0	0	0	14	0	0	0	29	0	5	2

Données relatives

IO	<i>% marques OPQ / taille totale pers.</i>
Fernande	1.090093
Lorraine	4.473928
Lucille	2.382926
Yvette	1.764095

*Encore une fois, si vous permettez (1998)***Données brutes**

EF	<i>a.</i>	<i>b.</i>	<i>c.</i>	<i>d.</i>	<i>e.</i>	<i>f.</i>	<i>g.</i>	<i>h.</i>	<i>i.</i>	<i>j.</i>	<i>k.</i>	<i>l.</i>
Nana	198	96	0	7	497	13	9	55	225	11	40	44
Le Narrateur	55	45	0	0	116	2	3	7	111	9	5	15
N1	0	0	0	0	0	0	0	0	0	0	1	0
N2	55	45	0	0	116	2	3	7	111	9	4	15

Données relatives

EF	<i>% marques OPQ / taille totale pers.</i>
Nana	10.26456
Le Narrateur	8.095029
N1	0.091912
N2	10.61307

Rem. : N1 et N2 sont deux sous-types du personnage Le Narrateur.

2.4. Valeur distinctive des variables

Ce que je cherche à tester

Les différences de proportion d'usage de phénomènes caractéristiques de l'OPQ observées pour chaque personnage, **à l'intérieur d'une pièce**, sont-elles significatives, ou sont-elles dues au hasard ?

Tests retenus

Test de Welch : mesure une différence de moyennes.

Test de Fisher : mesure ici une corrélation entre identité personnage et proportion des marques d'OPQ.

Les deux tests sont censés donner des résultats cohérents.

Tests compatibles avec le type des données recueillies

En particulier :

- On n'a aucune information permettant de faire l'hypothèse que la population de locuteurs, dont les groupes de personnages sont des échantillons, a une distribution normale.
- Les données sont proportionnelles,
- Leur distribution n'est pas symétrique,
- Il n'y a pas d'homogénéité des variances.

Il est dans ce cas-là conseillé d'utiliser **des tests de significativité non paramétriques** et de préférer le Fisher au Khi carré, plus « exigeant » et plus robuste. Cf. Rietveld & van Hout 2005 : 125, Manning & Schütze 2002 : 50-54, Ramousse et al. 1996 : chap. 3, Yaeger 2007 : 39-47.

Calculs faits avec le logiciel R

<http://cran.r-project.org/> (Crawley 2005, Vasishth 2005, Venables & Ripley 2002)

2.5. Résultats des tests pour *L'Impromptu d'Outremont*

[Exemplier D2, p. 7]

ANALYSE DE L'IMROMPTU D'OUTREMONT --- seuil = 5%

Test de Welch et test de Fisher

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]		Fernande	Lorraine	Lucille	Yvette
[2,]	Fernande		1	1	1
[3,]	Lorraine	1		1	1
[4,]	Lucille	1	1		
[5,]	Yvette	1	1		

Différence

1 significative

Ø non significative

Total occurrences marques OPQ / taille totale discours du personnage

Fernande : 0.010900093 (taille totale = 6238)

Lorraine : 0.04473928 (taille totale = 3241)

Lucille : 0.02382926 (taille totale = 4826)

Yvette : 0.01764095 (taille totale = 2891)

Commentaire des résultats pour *L'Impromptu*

Trois groupes de personnages se détachent :

1. Fernande, qui correspond au pôle – OPQ.
2. Lorraine, l'extrême inverse, qui correspond au pôle + OPQ.
3. Yvette et Lucille, qui sont associées à des profils linguistiques intermédiaires.

2.6. Résultats des tests pour *Encore une fois, si vous permettez*

[Exemplier F2, p. 11]

ANALYSE DE ENCORE UNE FOIS SI VOUS PERMETTEZ --- seuil = 5%

Test de Welch et test de Fisher

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]		Nana	Le Narrateur	N1	N2
[2,]	Nana		1	1	
[3,]	Le Narrateur	1		1	1
[4,]	N1	1	1		1
[5,]	N2		1	1	

Différence
1 significative
Ø non significative

Total occurrences marques OPQ / taille totale discours du personnage

Nana : 0.1026456 (taille totale = 11642)

Le Narrateur : 0.08095029 (taille totale = 4546)

N1 : 0.0009191176 (taille totale = 1088)

N2 : 0.1061307 (taille totale = 3458)

Commentaire des résultats pour *Encore une fois*

Structure plus complexe.

Deux personnages textuellement déclarés

1. Nana (une comédienne)
2. Le Narrateur (= N1 + N2) (un seul comédien)

N1 : Le Narrateur s'adresse au public.

N2 : Le Narrateur s'adresse à Nana.

Trois profils linguistiques différents

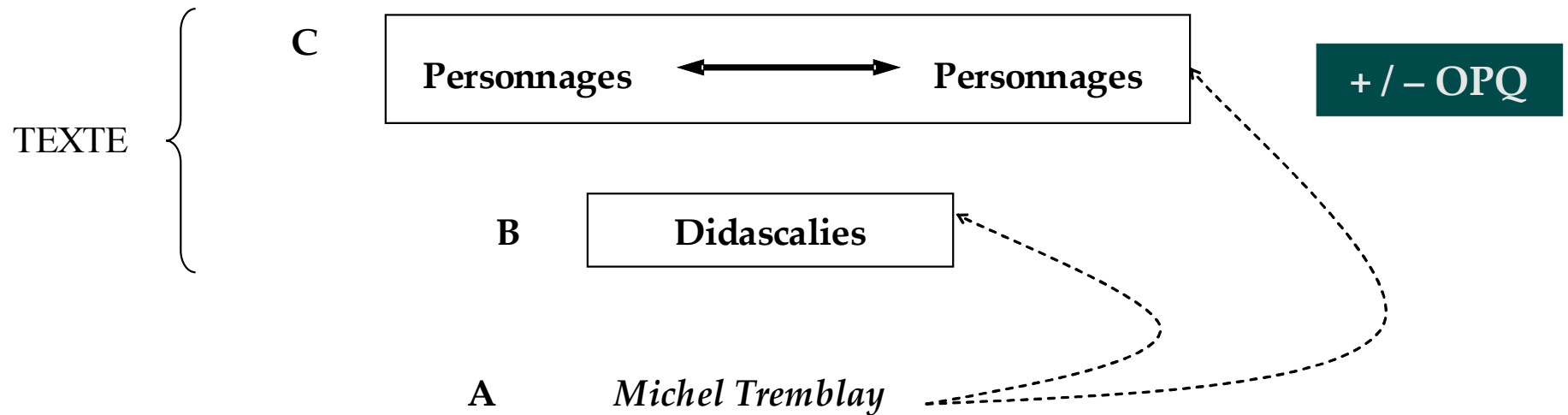
1. Nana et N2, qui se présentent comme le pôle + OPQ.
2. N1, qui correspond au pôle – OPQ.
3. Le Narrateur, ce collage de N1 et N2, dont le profil est hybride.



3. Interprétation (stylistique) des résultats

3.2. Profils et structure textuelle : *L'Impromptu...*

[Exemplier D3, p. 8]



Légende

A : l'écrivain

↔ relation locuteur/allocutaire

-----> est perçu comme l'origine de

Extraits de *L'Impromptu remontrera*

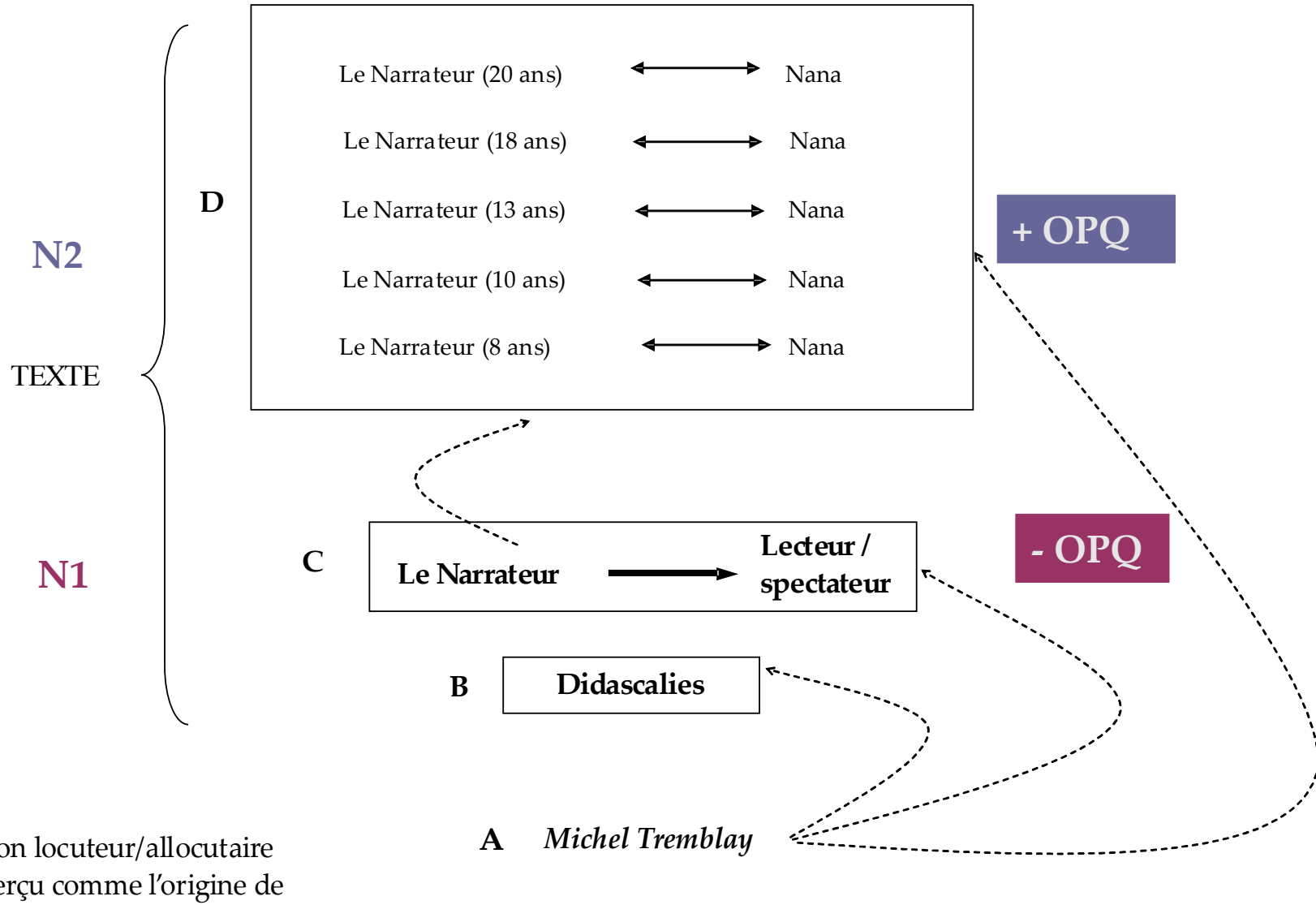
[Exemplier D4, p. 8]

LORRAINE. Chus tannée de la voir frémir pis trembler chaque fois que j'dis quequ'chose qui est pas vérifiable dans le dictionnaire. [...] Quand j'vois les sourcils y froncer, pis la bouche y durcir, pis le nez y pincer [...] j'ai envie de me sacrer à ses pieds en demandant pardon de l'avoir offensée, elle la vierge de la langue française ! C'est pas des farces, des fois, quand j'sors d'ici, j'fais attention comment j'parle. (IO, 67 sqq.)

FERNANDE. Le cri a remplacé la délicatesse dans ce pays ! Vous n'êtes pas fatiguées d'entendre crier au théâtre, au cinéma, à la télévision ? [...] N'est-il plus possible aujourd'hui, ici, maintenant, de dire les choses sans les hurler ? [...] Rien n'est assez vil [...] pour attirer le spectateur dans une salle [...] (IO, 97)

3.2. Profils et structure textuelle : *Encore une fois...*

[Exemplier F3, p. 12]



Rappel des groupes + / – OPQ pour *Encore une fois...*

Deux personnages textuellement déclarés

1. Nana (une comédienne)
2. Le Narrateur (= N1 + N2) (un seul comédien)

N1 : Le Narrateur s'adresse au public.

N2 : Le Narrateur s'adresse à Nana.

Trois profils linguistiques différents

1. Nana et N2, qui se présentent comme le pôle + OPQ.
2. N1, qui correspond au pôle – OPQ.
3. Le Narrateur, ce collage de N1 et N2, dont le profil est hybride.

Structure énonciative plus complexe

Strate fictionnelle et énonciative supplémentaire (D).

Le Narrateur est le pivot de deux mondes possibles et de deux natures d'interaction.

- *Niveau C* : image d'un dramaturge mature s'adressant au public en lui présentant le niveau D.
- *Niveau D* : Le Narrateur s'adresse à Nana. Rétrospectives successives où Le Narrateur a 8, 10, 13, 18 et 20 ans, mais où Nana paraît intemporelle.

Structure fictionnelle mise à nu

La différenciation des profils + OPQ et – OPQ n'est plus sociolinguistique, mais affective et métalittéraire : elle marque un voyage dans le temps de l'enfance et l'enchâssement de mondes fictionnels. Les marques d'oralité indiquent ici les frontières fictionnelles.

Extraits de *Encore une fois, si vous permettez*

[Exemplier F4, p. 12]

Le plateau est vide.

Le Narrateur entre, s'assoit sur une chaise qu'il ne quittera pas jusqu'à la fin. [...]

Nana, elle, envahit le plateau aussitôt arrivée, l'habite, le domine, en fait son royaume.

N1 LE NARRATEUR. Ce soir, personne ne viendra crier : « Pour qui sont ces serpents qui sifflent sur nos têtes ? », ni murmurer : « Va, je ne te hais point » en se tordant les mains. [...] Ce que vous verrez, ce sera une femme toute simple, une simple femme qui viendra vous parler [...] (EF, 9-10)

~

N2 LE NARRATEUR. J'ai pas fait exiprès.

NANA. Comment ça, t'as pas fait exiprès ! T'as pitché un morceau de glace en dessous d'une voiture en marche [...] (EF, 11)

~

N2 LE NARRATEUR. Chus pus sûr si j'ai ben ben envie de te conter tout ça.

NANA. Moi, j'ai le goût ! Envoye ! (EF, 15)



4. Conclusions et prolongements

4.1 . Conclusions de l'analyse

4.1.1. Évolution des fonctions des marques d'OPQ dans le corpus *tremblay*

Ce qu'on observe avec ces deux pièces, se vérifie pour le reste du corpus et pour l'ensemble de l'oeuvre de M. Tremblay (cf. Dargnat 2006).

- Passage d'une *satire sociale de la langue* du Montréal francophone à une *autobiographie affective* en hommage à sa mère (= notion de « langue maternelle » change de référent).
- Autoréférentialité différente. Dans *L'Impromptu*, Tremblay fait référence à lui-même à travers la mention d'une de ses pièces, dans *Encore une fois*, il se fait référence à lui-même comme dramaturge en train de créer une partie de la pièce représentée.

4.1.2. Évolution qui correspond en partie à l'évolution des modèles esthétiques québécois

- Du *Nous collectif* au *Je introspectif* (Cambron 1989).
 - De *l'esthétique de la fondation* et de *la transgression* à *l'esthétique de la ritualisation* (Gauvin 2002, Nepveu 1999).
 - Redéfinition des frontières génériques notamment par les procédés d'*autoréférentialité*, de *mise en abyme* et de « *métalepse narrative* ».
- Cf. Dion et al. 2001 pour la littérature québécoise, Genette 2004, Pier & Schaeffer 2005, Ryan 1991 pour des réflexions générales.

4.2 . Prolongements, en cours ou à faire... un jour

4.2.1. Fonctionnalités du logiciel *Weblex*

Comparaison des résultats obtenus avec par exemple le calcul des *spécificités* (Lafon 1984) de chaque variable pour les personnages de chaque pièce. Mais :

- Tâche rendue difficile par la nature de certaines variables (décompte semi-manuel / relevé automatique des occurrences dans *Weblex*),
- Validité statistique de la procédure à déterminer par rapport aux hypothèses et aux données,
- Logiciel et fonctionnalités en cours de réévaluation (projet ANR « Fédération des recherches et développements en textométrie autour de la création d'une plate-forme logicielle ouverte », porté par S. Heiden, 2007-2009).

4.2.2. Réduction de la liste de variables

Existe-il un nombre restreint de phénomènes linguistiques qui suffiraient à déterminer les profils + / – OPQ des personnages ? Les douze variables sont-elles toutes pertinentes, sont-elles redondantes ?

- Pour les personnages d'une même pièce ?
- Pour tous les personnages ?

Rem. : une **analyse factorielle** demande en théorie qu'on puisse faire l'hypothèse d'une distribution normale (Gorsuch 1983 (chap. 8) : 148). Ce qui n'est pas le cas.

4.2.3. Autre utilisation des tests de *Welch* et de *Fischer*

En fusionnant les cinq pièces (= 37 personnages)

Ne change pas les résultats obtenus localement (dans une pièce).

Ajoute des comparaisons entre les personnages de pièces différentes :

- un personnage avec une valeur + OPQ dans « sa » pièce, peut ressembler à un personnage avec une valeur – OPQ dans une autre pièce,
- déconstruction de l'effet cognitif dû à l'environnement d'un univers fictionnel clos.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
"1"	-	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	
"2"	-	-	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	
"3"	-	-	-	0	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
"4"	-	-	-	-	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
"5"	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	
"6"	-	-	-	-	-	-	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	
"7"	-	-	-	-	-	-	-	0	0	0	0	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
"8"	-	-	-	-	-	-	-	-	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	0
"9"	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	0	1	1	0
"10"	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	0	1	1	0
"11"	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	1	0	0	0	0	1	0	1	1	1	1	1	0	1	0	0	0	0	0	1	1	0
"12"	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	
"13"	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0
"14"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	
"15"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
"16"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
"17"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	
"18"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0
"19"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0
"20"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	1	1	1	1	1	0	1	0	1	1	1	1	
"21"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	
"22"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	0	1	0	1	1	1	1	
"23"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	
"24"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	
"25"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	1	1	1	1	1	1	1	
"26"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	
"27"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	1	0	1	0	1	1
"28"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	0	1	0	1	1
"29"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	0	1	0	1	1
"30"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	0	1	1	0
"31"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	1	1	0
"32"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0	1	1	0
"33"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	0	0
"34"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	0
"35"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1
"36"	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

4.2.4. Agrégation (*clustering*) des individus (= les personnages)

Démarche exploratoire au moyen d'algorithmes assez standard :

DIANA (*D*ivisive *A*NALysis)

AGNES (*A*Gglomerative *N*ESTing)

PAM (*P*artitioning *A*round *M*edoid)

Rem. Calculs non terminés, en cours d'interprétation.

Démarche bien documentée (cf. Kaufman & Rousseeuw 1990).

4.2.5. Perception des différences et « stylistique cognitive »

cf. Semino & Culpeper 2002

– Tester la **pertinence des douze variables**, leur « rentabilité stylistique » sur des sujets-lecteurs. L'effet OPQ n'est pas qu'une réalité statistique mais aussi le résultat d'un **processus cognitif plus ou moins individuel de catégorisation** (cf. Clopper & Pisoni 2005).

– Tester la **différence entre des classements objectifs « aveugles » basés sur la statistique et les classements intuitifs** à la simple lecture des textes/extraits de textes.

Demande un protocole expérimental assez lourd, et à déterminer.



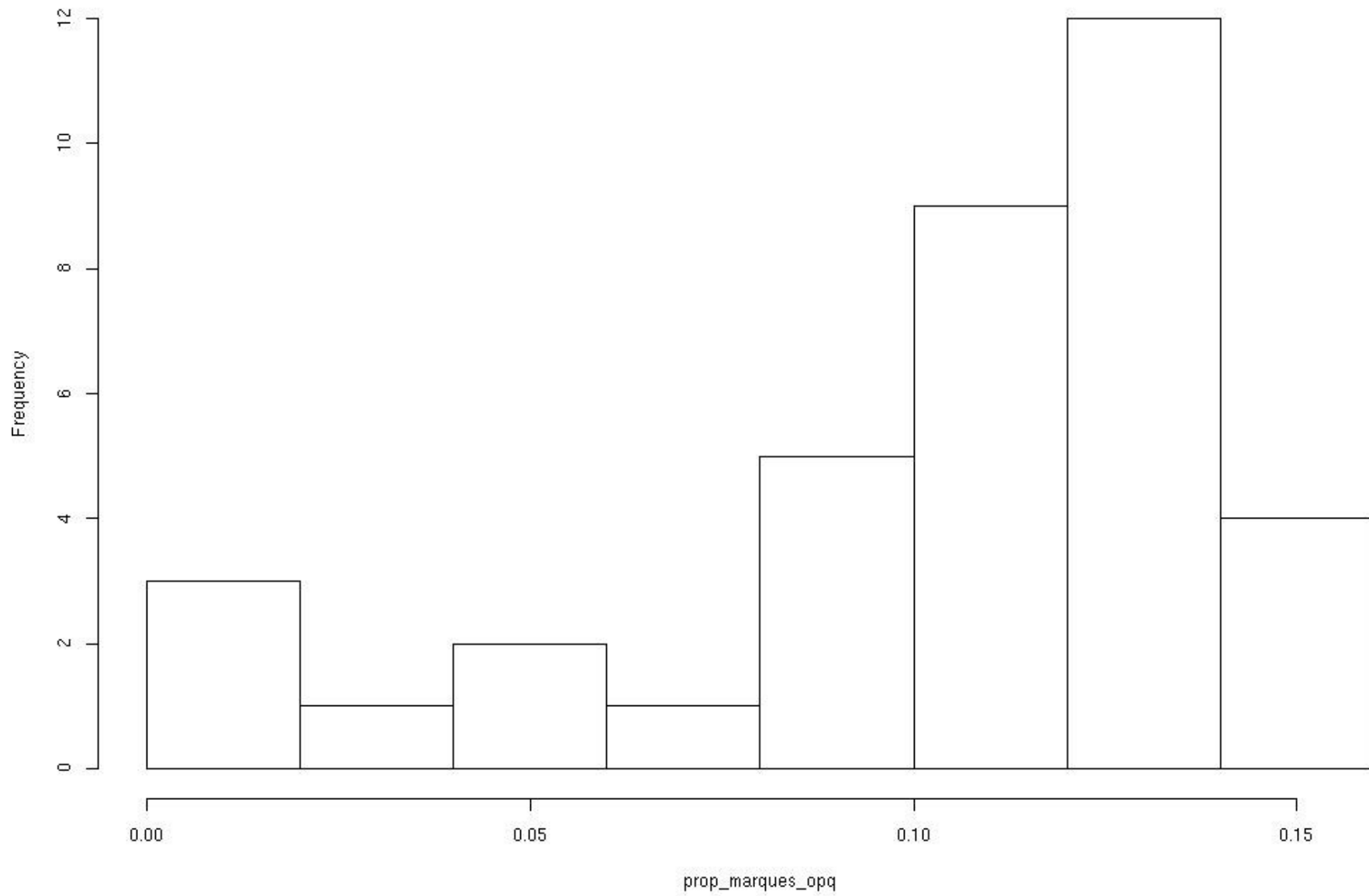
Villa R
Paul Klee, 1919.

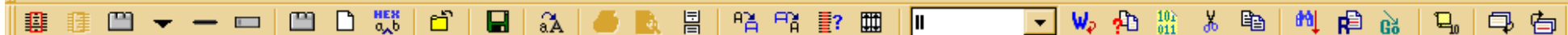
Je vous remercie de votre attention.

Courriel : mathilde.dargnat@free.fr

Page perso. : <http://mathilde.dargnat.free.fr>

Histogram of prop_marques_opq





```
#La variable du pis et des apostrophes.: pas d'homogénéité des variances + dissymétrie
```

```
#variance
```

```
> var(tremblay_data[1:10,2])
[1] 0.1390354
> var(tremblay_data[11:20,2])
[1] 0.2993156
> var(tremblay_data[21:30,2])
[1] 0.6094156
> var(tremblay_data[31:37,2])
[1] 0.4622593
> var(tremblay_data[1:10,6])
[1] 1.262937
> var(tremblay_data[11:20,6])
[1] 1.363870
> var(tremblay_data[21:30,6])
[1] 4.112069
> var(tremblay_data[31:37,6])
[1] 2.890983
```

```
#coef. de symétrie
```

```
> library(e1071)
Le chargement a nécessité le package : class
> skewness(tremblay_data[1:10,2])
[1] -0.4697702
> skewness(tremblay_data[10:20,2])
[1] -0.261223
```

```
# Problèmes identiques sur l'ensemble des variables
```

```
> tremblay_table_fisher <- read.table("/home/Mathilde/tremblay_data_fisher.txt",header=T,row.names=1)
> tremblay_table_fisher
      personnages marques_opq autres
gl    germaine_lauzon      484  3213
ll    linda_lauzon        130   798
ro    rose_ouimet        498  3074
gj    gabrielle_jodoin    136   820
ldc   lisette_de_courval   54   998
```



```

mb      ma_brouillette      118   757
yl      yvette_longpre      54    532
dnv    des_neiges_verrette  131   1113
td      therese_dubuc       137   983
as      angeline_sauve       170   1228
rb      rheauna_bibeau      106   850
lp      lise_paquette        110   705
gm      ginette_menard     24    187
pg      pierrette_guerin    134   825
lu      lucienne           370   2348
den     denise                 276   1690
mon     monique               236   1491
nic     nicole                 155   1215
alb     albertine              166   1370
ch      charlotte              143   899
gab     gabriel                 309   1818
ser     serge                  413   2714
fer     fernande                68    6170
lor     lorraine               145   3096
lle     lucille                115   4711
yv     yvette                 51    2840
cl     claude                 314   3616
m1     madeleine1            312   3253
m2     madeleine2            284   3008
al1     alex1                 394   3034
al2     alex2                 279   2305
ma1     mariette1            133   970
ma2     mariette2            118   1171
na     nana                  1195  10447
nr     narrateur             368   4178
nr1    narrateur1            1    1087
nr2    narrateur2            367   3091

```

```

> marques_opq <- c(tremblay_table_fisher[,2])
> marques_opq
 [1] 484 130 498 136 54 118 54 131 137 170 106 110 24 134 370
[16] 276 236 155 166 143 309 413 68 145 115 51 314 312 284 394
[31] 279 133 118 1195 368 1 367
> marques_autres <- c(tremblay_table_fisher[,3])

```



```
> marques_autres <- c(tremblay_table_fisher[,3])
> prop_marques_opq <- rep(1,37)
> for (i in 1:37) {prop_marques_opq[i] <- marques_opq[i]/(marques_opq[i]+marques_autres[i])}
> prop_marques_opq
 [1] 0.1309169597 0.1400862069 0.1394176932 0.1422594142 0.0513307985
 [6] 0.1348571429 0.0921501706 0.1053054662 0.1223214286 0.1216022890
[11] 0.1108786611 0.1349693252 0.1137440758 0.1397288843 0.1361295070
[16] 0.1403865717 0.1366531558 0.1131386861 0.1080729167 0.1372360845
[21] 0.1452750353 0.1320754717 0.0109009298 0.0447392780 0.0238292582
[26] 0.0176409547 0.0798982188 0.0875175316 0.0862697448 0.1149358226
[31] 0.1079721362 0.1205802357 0.0915438324 0.1026455935 0.0809502860
[36] 0.0009191176 0.1061307114
> hist(prop_marques_opq)
> var(prop_marques_opq)
 [1] 0.001601835
> var(prop_marques_opq[1:18])
 [1] 0.0005264562
> var(prop_marques_opq[18:37])
 [1] 0.001889242
>
```