Table des matières des annexes

Annexe des figures

Les figures ci-dessous présentent respectivement l'exactitude de prédiction et la mesure F1 de chacun des deux classifieurs basée sur les corpus utilisés dans les expérimentations menées dans le chapitre 4 paragraphe § 4.8.

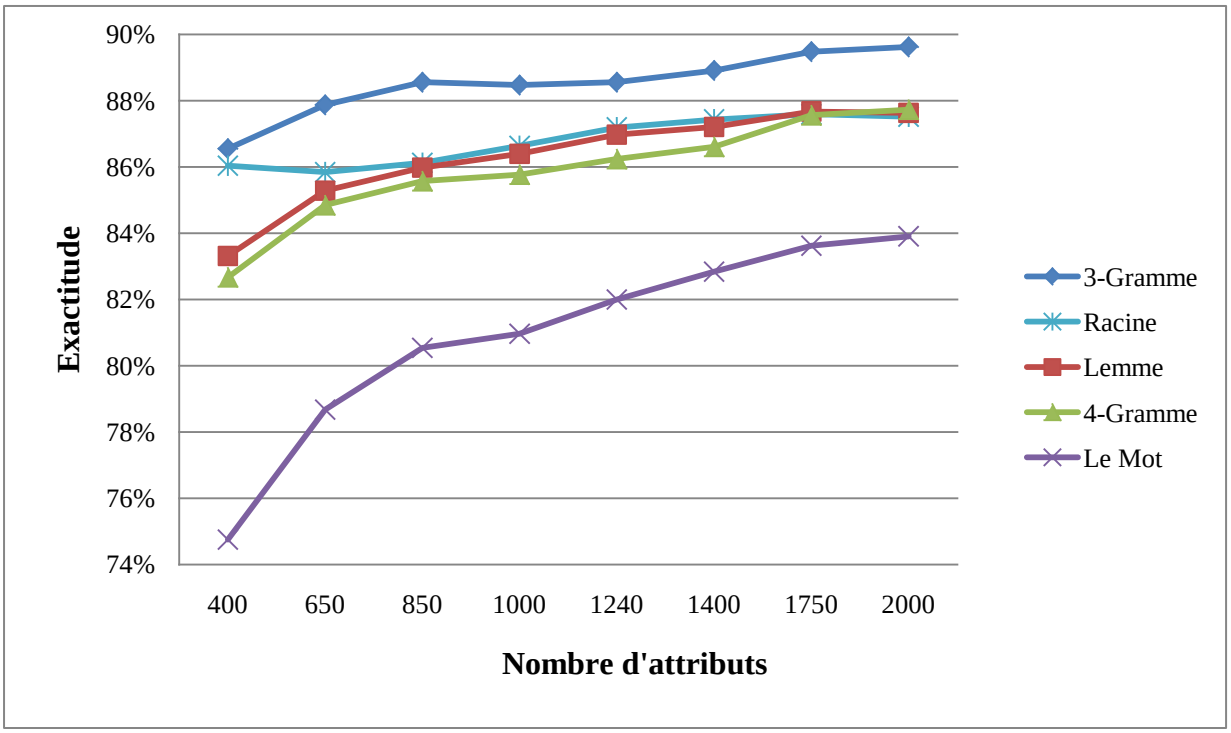

Figure 4.1 Exactitude du NBM avec GI

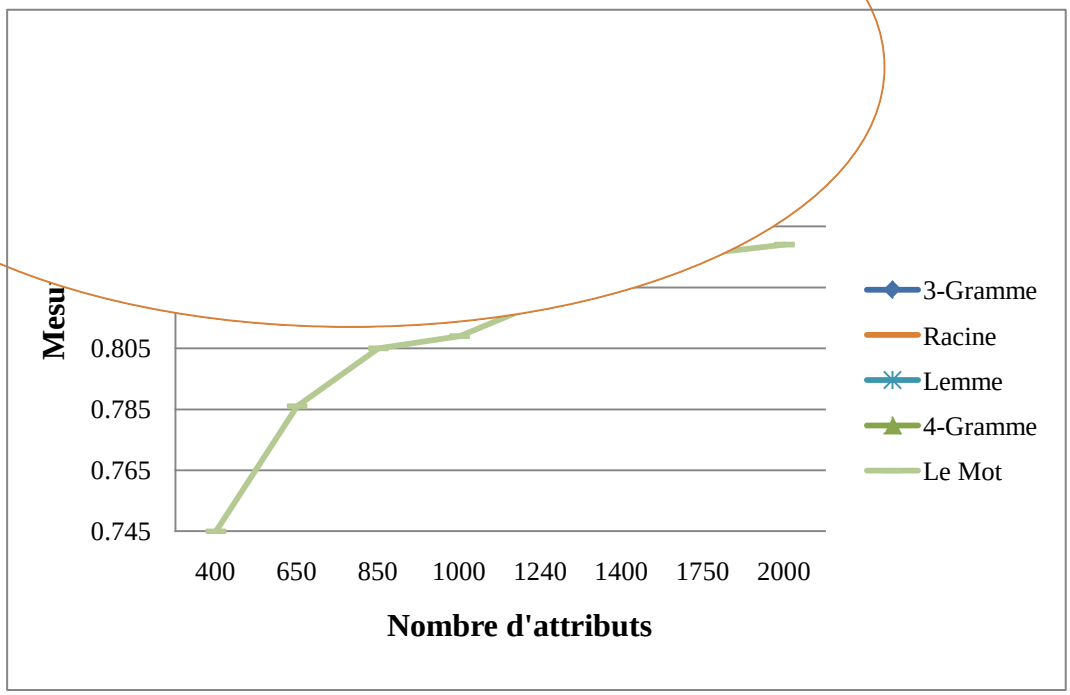

Figure 4.2. F1 du NBM avec GI

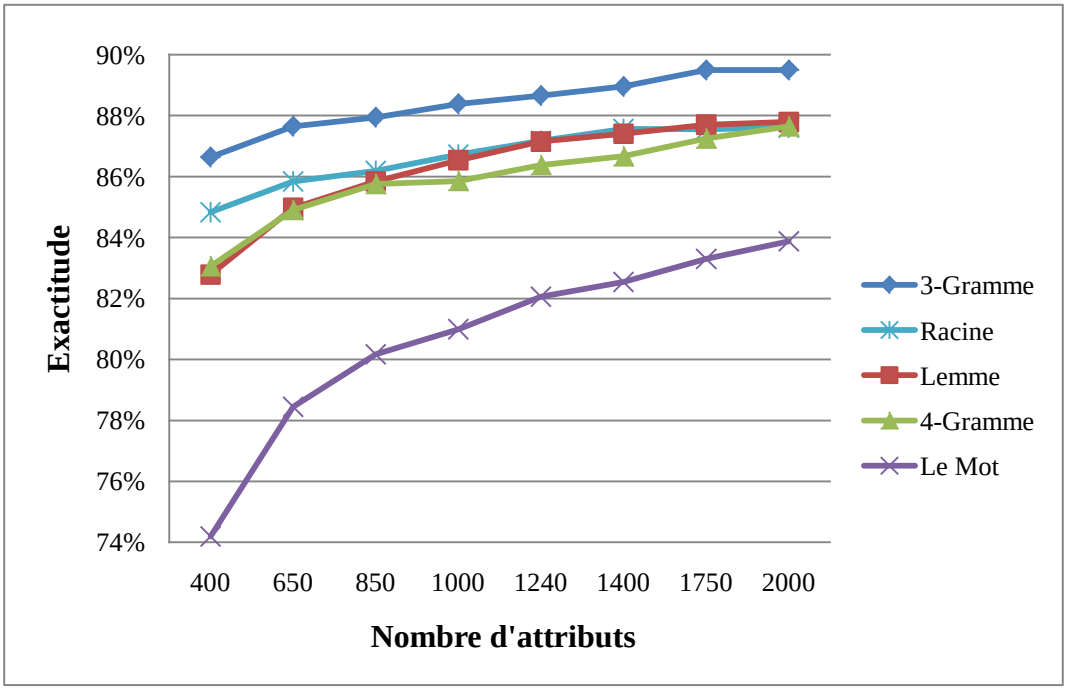

Figure 4.3. Exactitude du NBM et χ^2

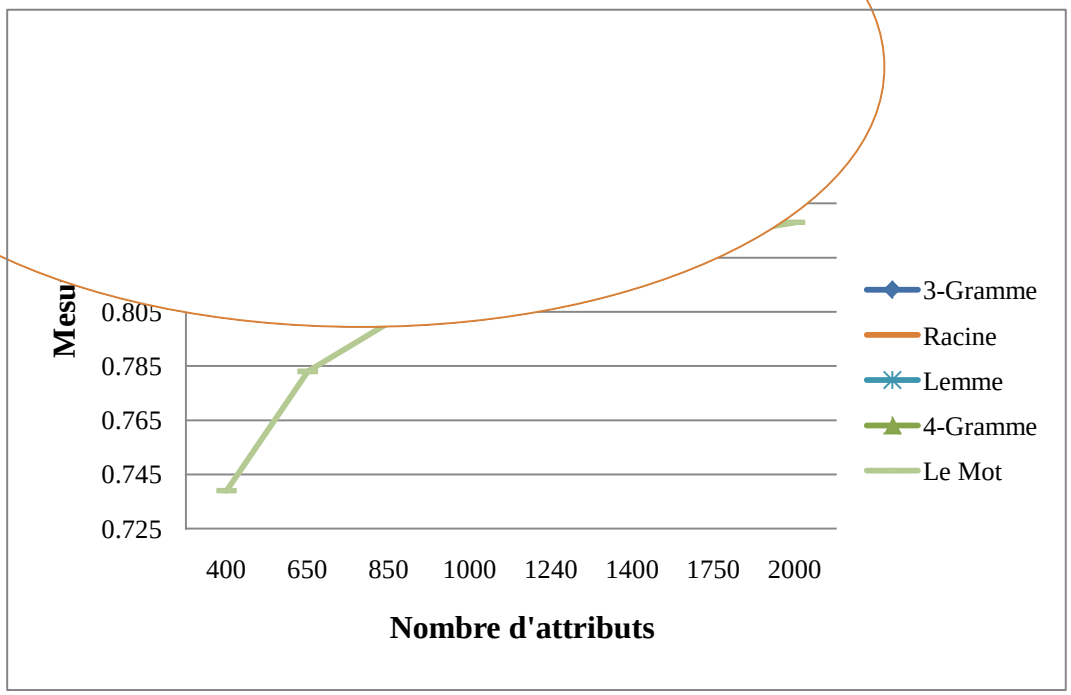

Figure 4.4. F1 du NBM et χ^2

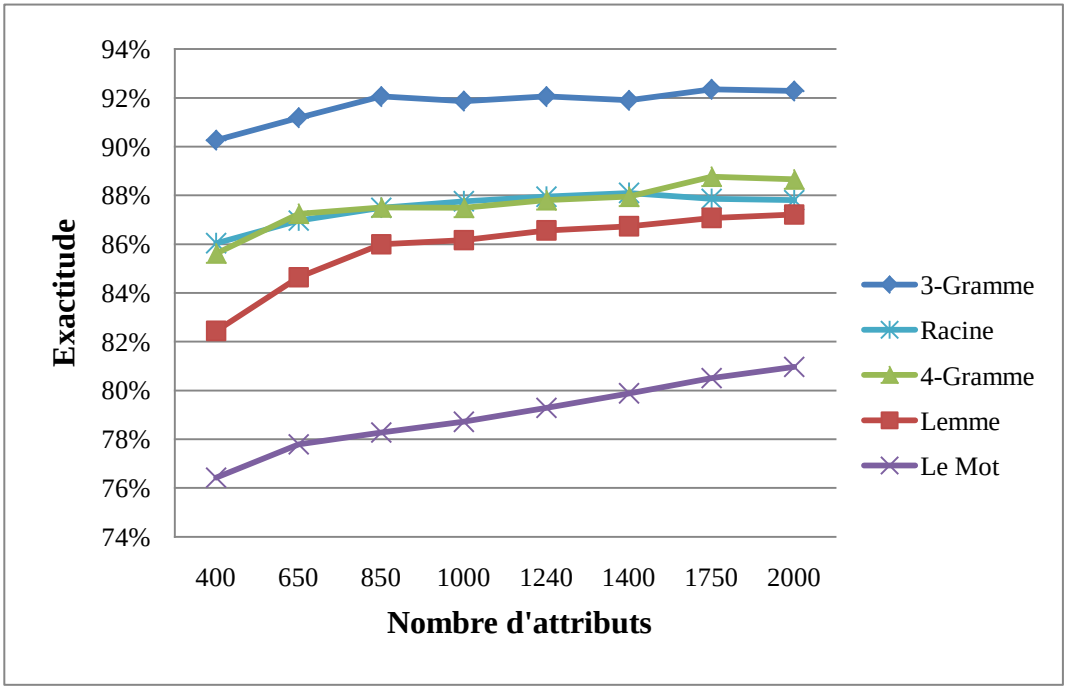

Figure 4.5. L'exactitude du SVM avec GI

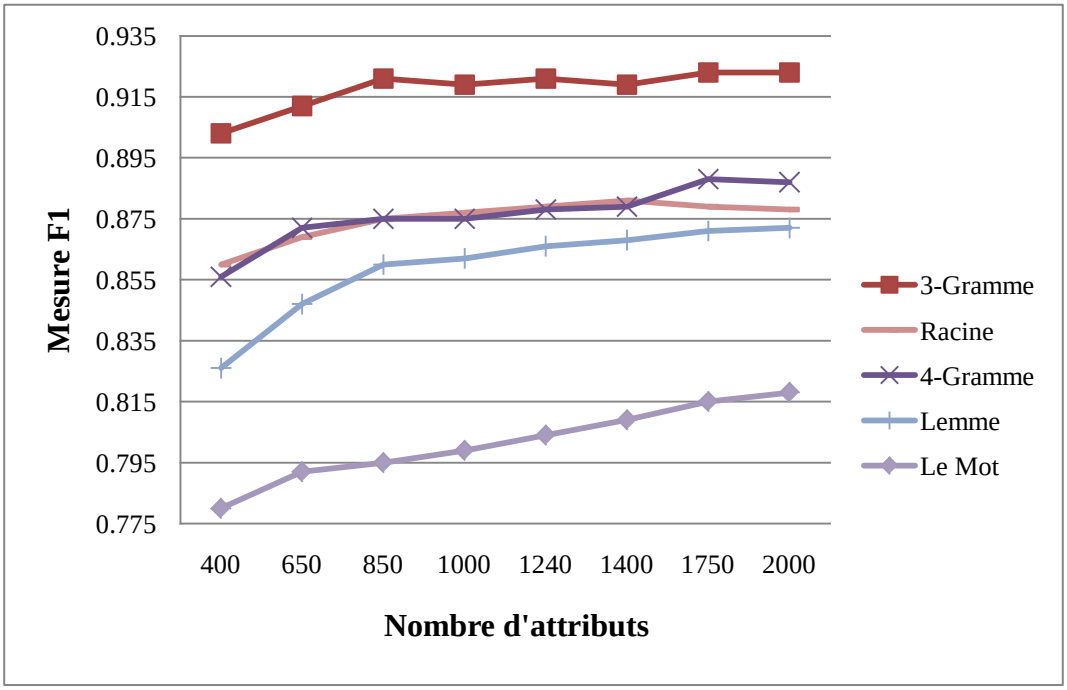

Figure 4.6. F-1 du SVM avec GI

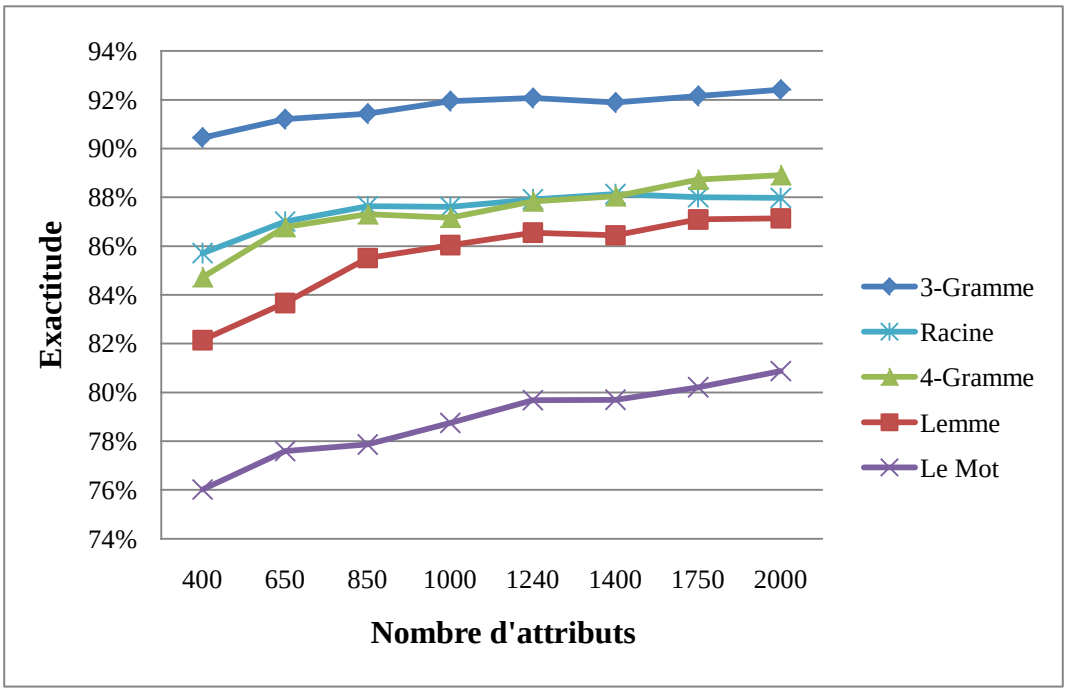

Figure 4.7. L'exactitude du SVM avec χ^2

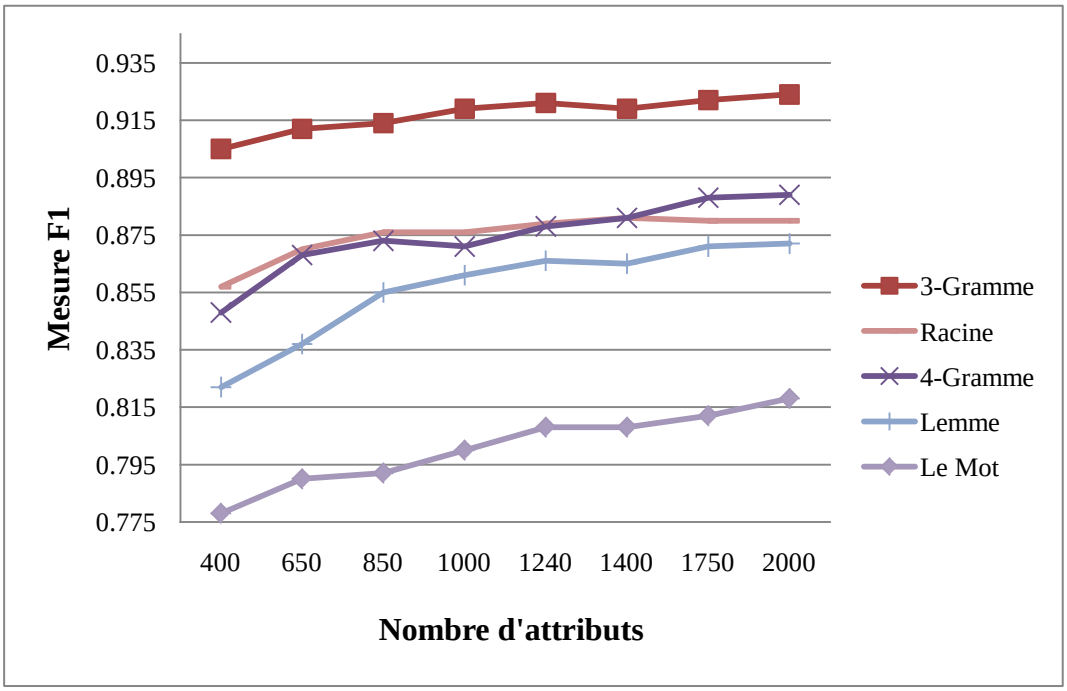

Figure 4.8. F-1 du SVM et χ^2

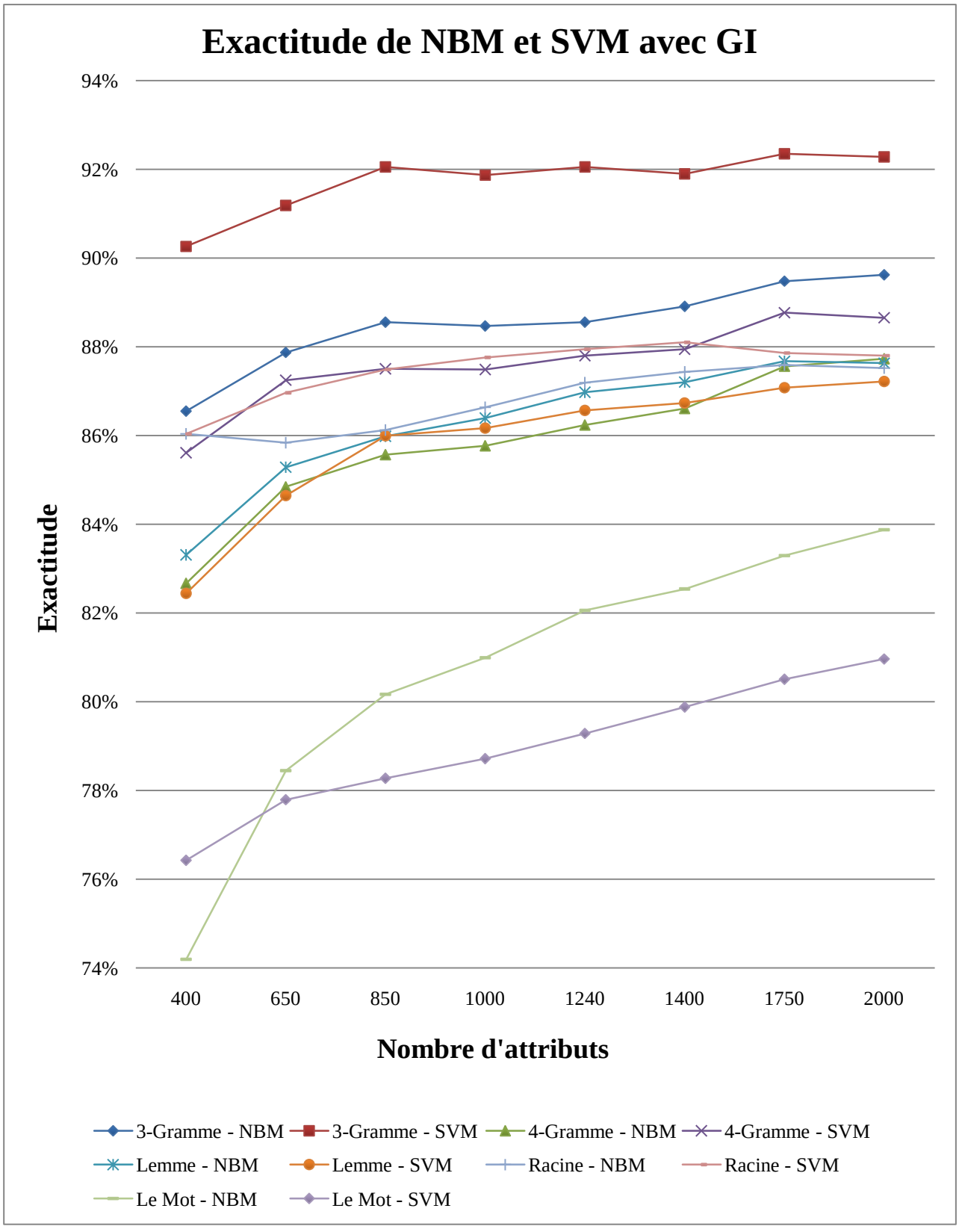

Figure 6.9 Exactitude de NBM et SVM avec GI

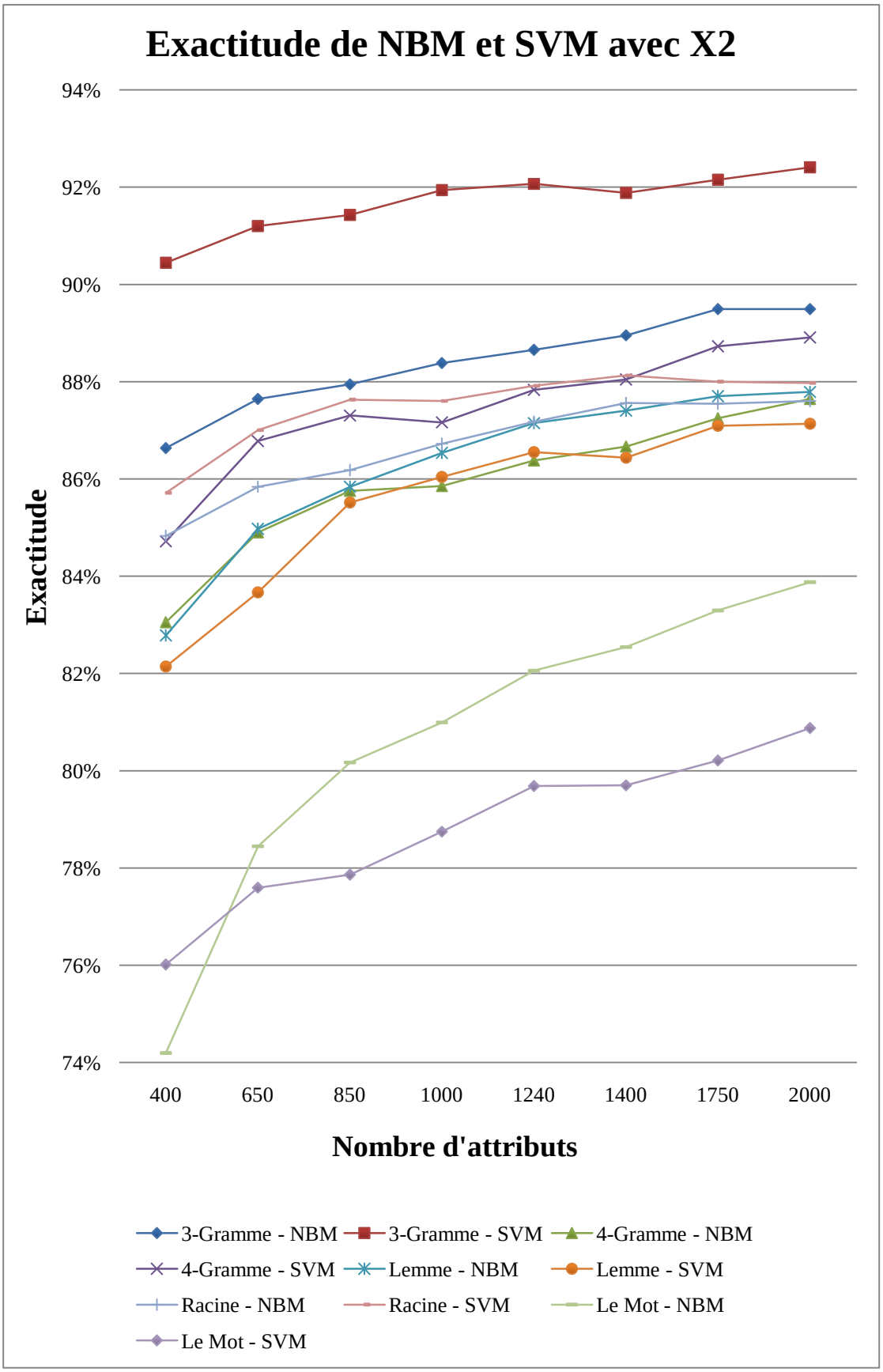

Figure 6.10 Exactitude de NBM et SVM avec χ^2

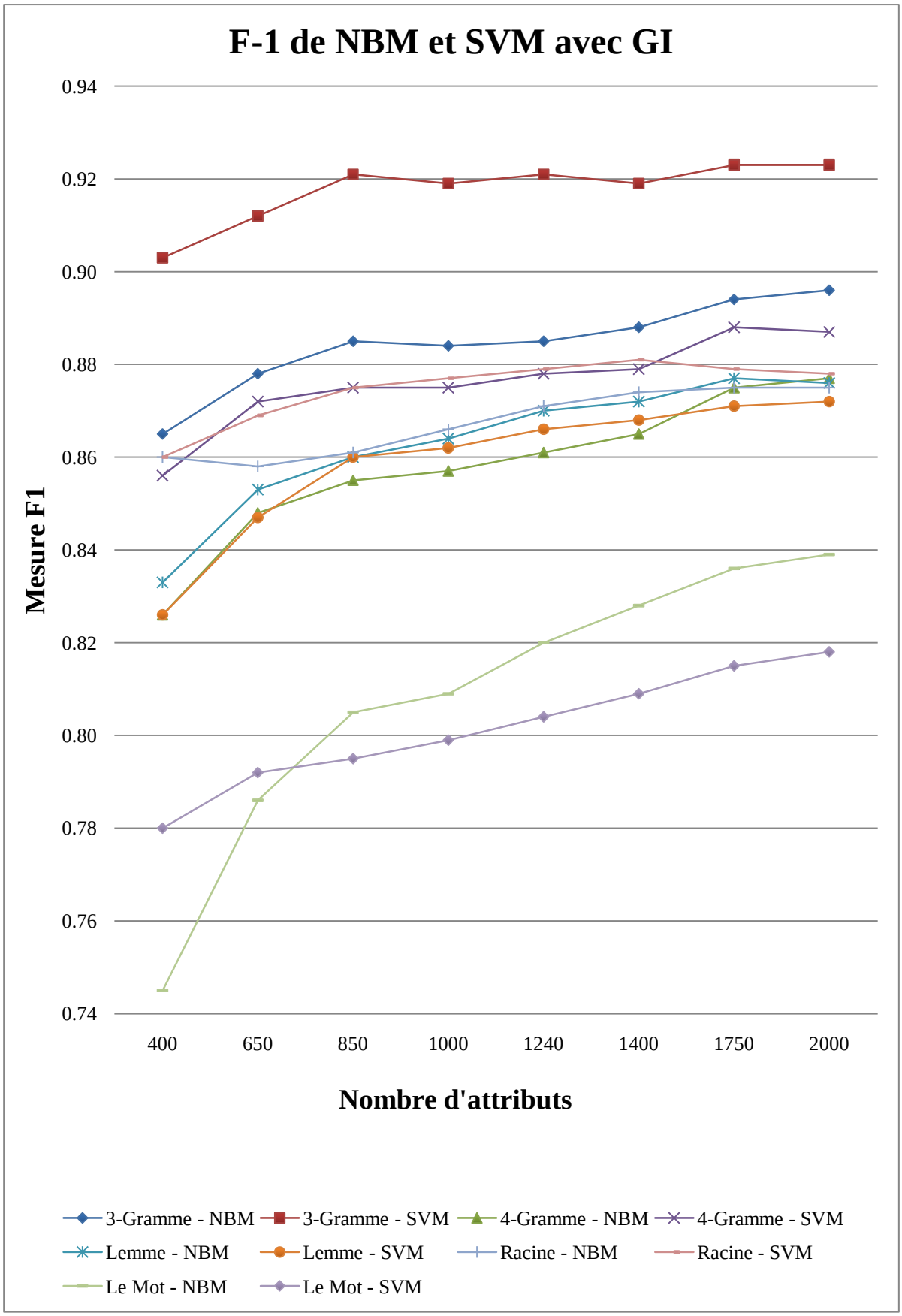

Figure 6.11 Mesure F-1 de NBM et SVM avec GI

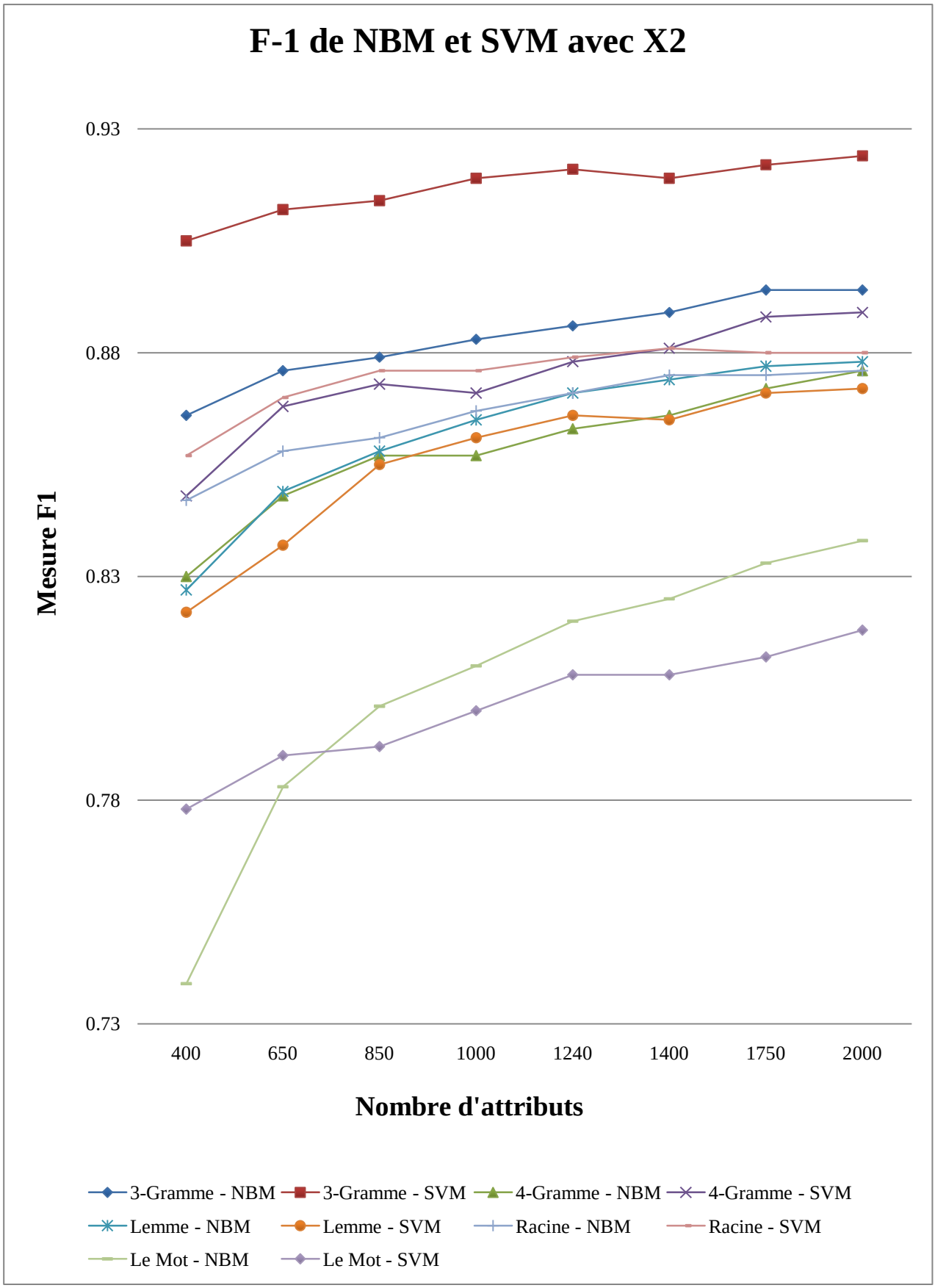

Figure 6.12 Mesure F-1 de NBM et SVM avec χ^2

Les figures ci-dessous présentent respectivement les mesures d'évaluation (rappel, précision, et F1) de chacun des classifieurs basé sur les corpus utilisés dans les expérimentations menées dans le chapitre 5 paragraphe §5.3.

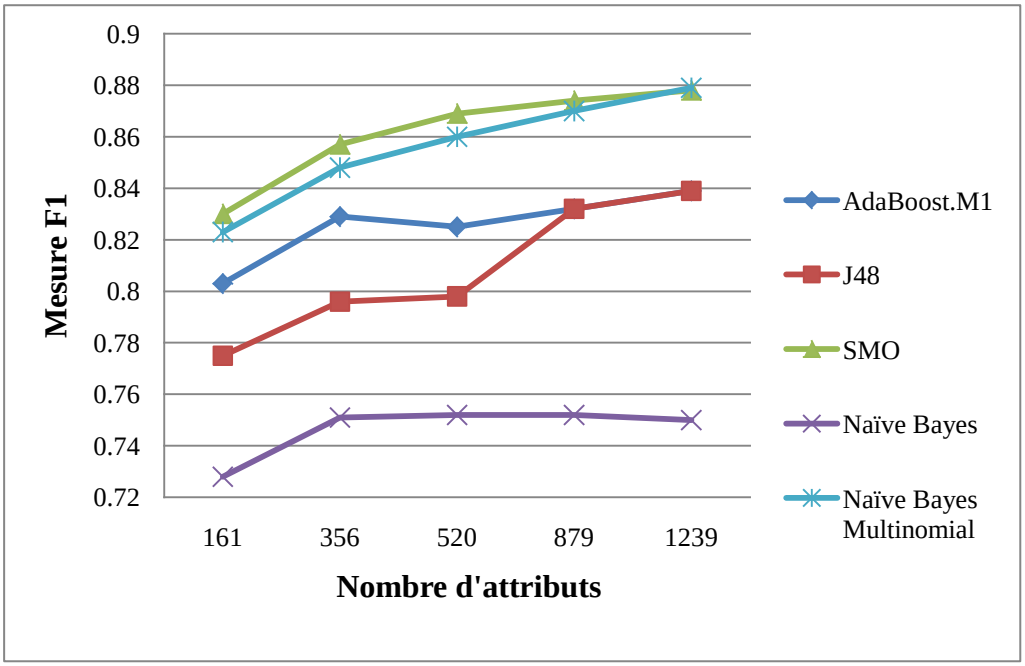

Figure 5.4 F1 en utilisant GI

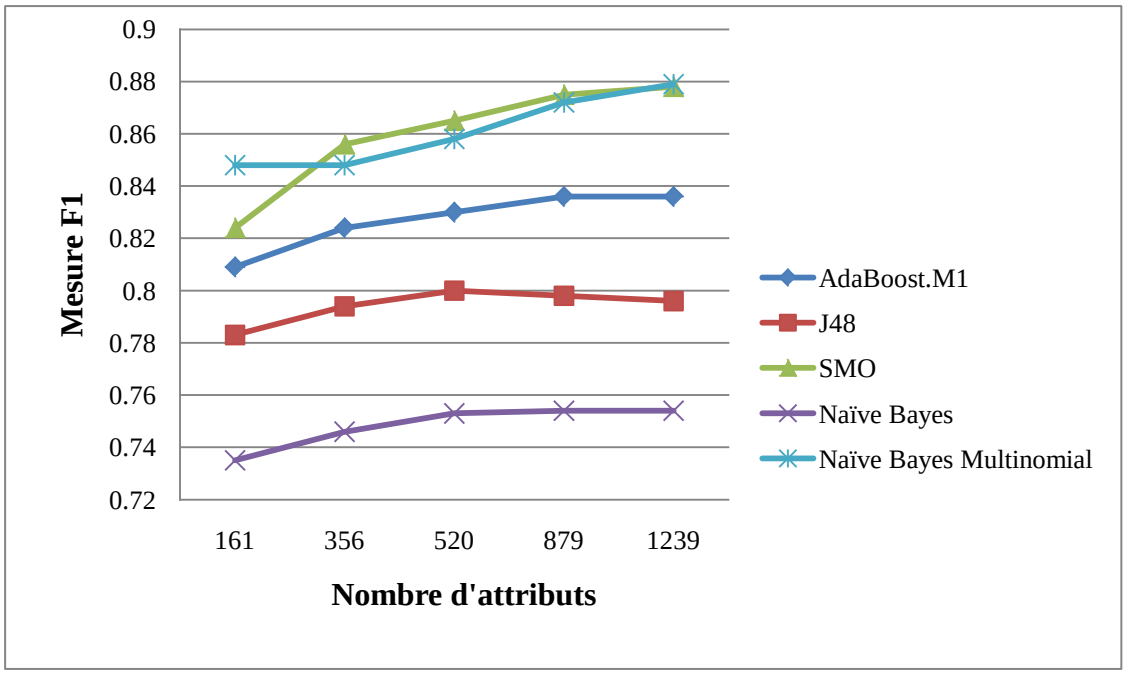

Figure 5.5 F1 en utilisant X^2

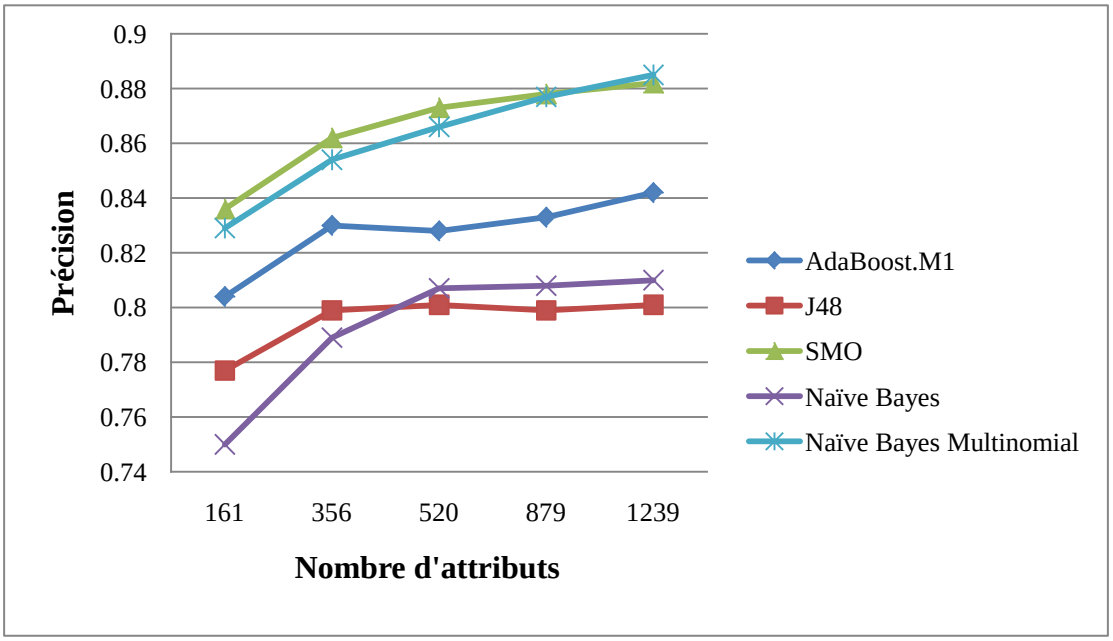

Figure 5.6 Précision en utilisant GI

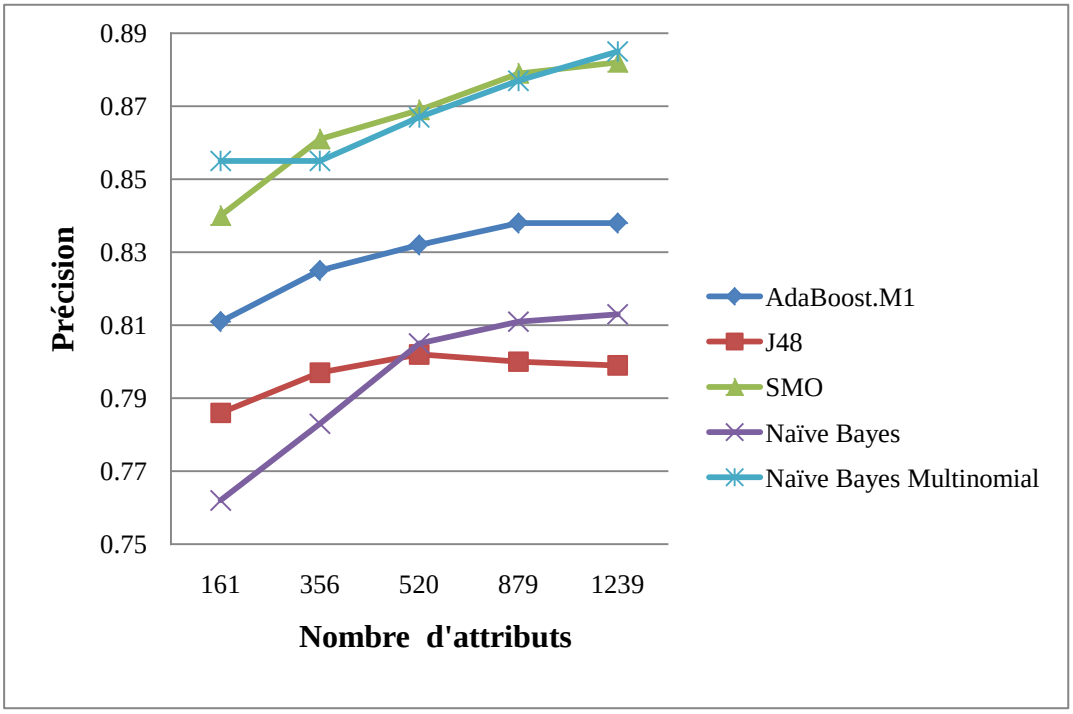

Figure 5.7 Précision en utilisant χ^2

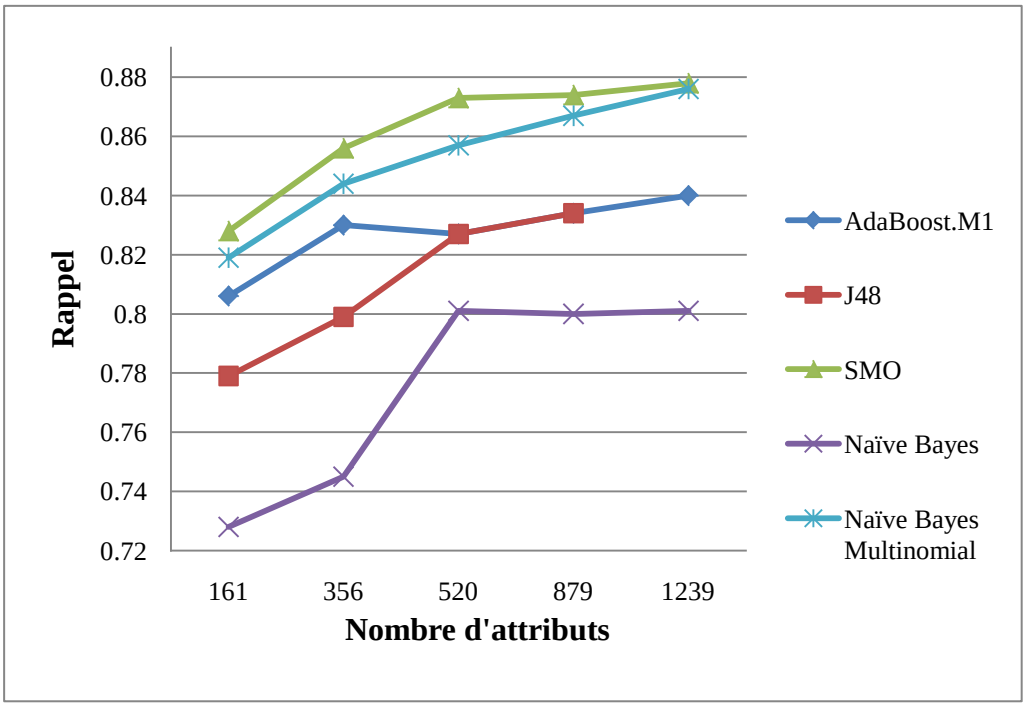

Figure 5.8 Rappel en utilisant GI

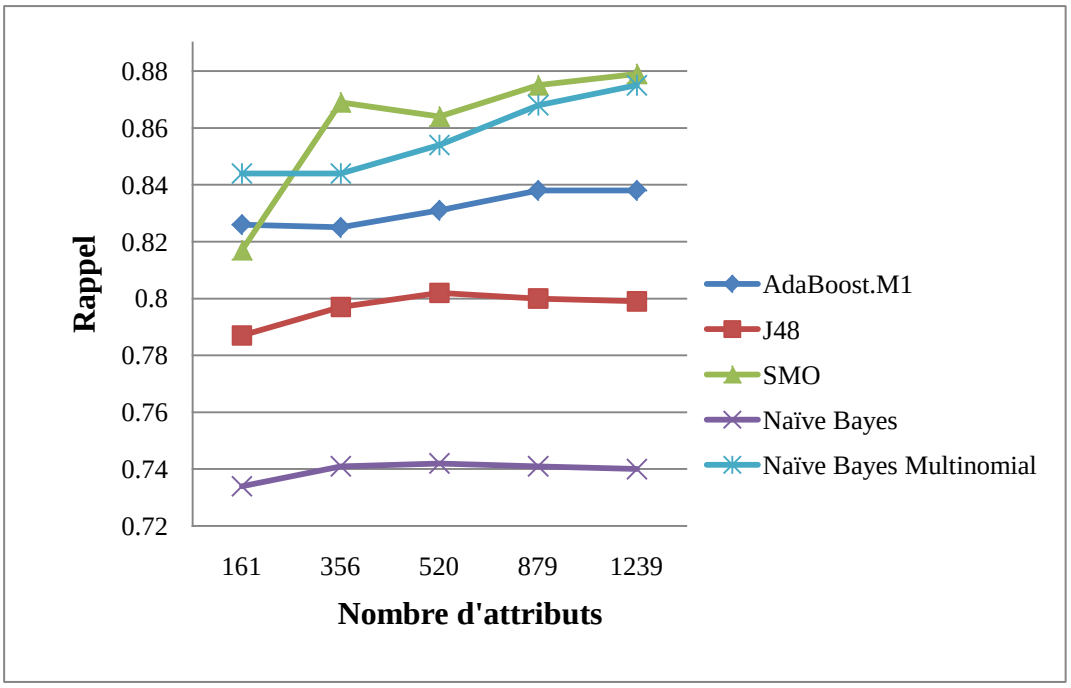

Figure 5.9 Rappel en utilisant χ^2

Annexe des tableaux

Les tableaux suivants présentent les résultats des expérimentations menées dans le chapitre 6 (cf. §6.6).

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**GI-SC**	**15**	**83.2954**	**0.832**	**1.6754**
	2	**RDFC**	**15**	**83.1248**	**0.83**	**1.6717**
	3	**3C-SC**	**15**	**83.1106**	**0.83**	**1.6715**
	4	GI		83.4518	0.834	1.6685
	5	FD		83.0822	0.83	1.6608
	6	**3C**		**82.9258**	**0.828**	**1.6572**
	7	$\boldsymbol{\chi^2}$		82.8121	0.825	1.6531
	8	$\boldsymbol{\chi^2}$**-SC**	**15**	**82.3145**	**0.819**	**1.6526**
	9	RC	6	64.4441	0.625	1.2736
	10	RG	21	34.3617	0.314	0.6723
	11	IM		18.9792	0.085	0.2747
SVM	1	**3C-SC**	**15**	**83.2243**	**0.832**	**1.6747**
	2	**RDFC**	**15**	**83.139**	**0.831**	**1.6728**
	3	GI		83.6224	0.836	1.6722
	4	**GI-SC**	**15**	**82.2101**	**0.832**	**1.6645**
	5	**3C**		**83.1817**	**0.832**	**1.6638**
	6	FD		83.1106	0.831	1.6621
	7	$\boldsymbol{\chi^2}$		82.3429	0.824	1.6474
	8	$\boldsymbol{\chi^2}$**-SC**	**15**	**81.76**	**0.818**	**1.6460**
	9	RC	6	64.3873	0.638	1.2860
	10	RG	21	32.6699	0.291	0.6323
	11	IM		19.3062	0.09	0.2830

Tableau 6.7 Résultats avec un corpus de 250 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**3C-SC**	**50**	**85.2573**	**0.851**	**1.7540**
	2	**GI-SC**	**50**	**85.1862**	**0.851**	**1.7540**
	3	**RDFC**	**50**	**85.2431**	**0.851**	**1.7493**
	4	**χ^2-SC**	**50**	**85.0441**	**0.849**	**1.7493**
	5	GI		85.3284	0.853	1.7183
	6	**χ^2**		85.0014	0.849	1.7142
	7	**3C**		**85.2858**	**0.852**	**1.7126**
	8	FD		85.2858	0.852	1.7126
	9	RC	30	71.0549	0.703	1.5529
	10	RG	21	35.2289	0.328	0.6907
	11	IM		20.3867	0.105	0.3872
SVM	1	**RDFC**	**50**	**86.2241**	**0.862**	**1.7791**
	2	**GI-SC**	**50**	**86.1814**	**0.861**	**1.7735**
	3	**3C-SC**	**50**	**86.2241**	**0.862**	**1.7730**
	4	**χ^2-SC**	**50**	**85.5274**	**0.855**	**1.7727**
	5	**3C**		**86.4657**	**0.864**	**1.7420**
	6	**χ^2**		85.9397	0.859	1.7406
	7	FD		86.3236	0.863	1.7346
	8	GI		86.1103	0.861	1.7345
	9	RC	3	69.9318	0.705	1.4993
	10	RG	21	33.6935	0.304	0.6544
	11	IM		21.1686	0.114	0.3942

Tableau 6.8 Résultats avec un corpus de 500 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**χ^2-SC**	**50**	**86.0108**	**0.859**	**1.7540**
	2	**3C-SC**	**50**	**85.7549**	**0.857**	**1.7495**
	3	**RDFC**	**50**	**85.7407**	**0.857**	**1.7493**
	4	**GI-SC**	**50**	**85.7265**	**0.857**	**1.7492**
	5	**3C**	**50**	**85.7123**	**0.857**	**1.7490**
	6	FD	21	85.6696	0.856	1.7273
	7	**χ^2**	2	85.9397	0.859	1.7197
	8	GI		85.698	0.856	1.7129
	9	RC	77	76.7984	0.764	1.5858
	10	RG	21	34.9019	0.327	0.6760
	11	IM		23.9266	0.148	0.3872
SVM	1	**3C**	**50**	**87.2192**	**0.872**	**1.7791**
	2	**GI-SC**	**50**	**86.9633**	**0.869**	**1.7735**
	3	**RDFC**	**50**	**86.9065**	**0.869**	**1.7730**
	4	**3C-SC**	**50**	**86.878**	**0.869**	**1.7727**
	5	**χ^2-SC**	**50**	**86.7501**	**0.867**	**1.7694**
	6	FD		87.1055	0.871	1.7420
	7	GI		87.0628	0.87	1.7406
	8	**χ^2**	2	86.7643	0.867	1.7360
	9	RC	77	74.0262	0.757	1.5511
	10	RG	21	33.5798	0.304	0.6544
	11	IM		24.4242	0.15	0.3942

Tableau 6.9 Résultats avec un corpus de 750 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**χ^2-SC**	**75**	**86.4231**	**0.863**	**1.7796**
	2	**RDFC**	**75**	**86.2951**	**0.862**	**1.7773**
	3	**3C-SC**	**75**	**86.2667**	**0.862**	**1.7771**
	4	**GI-SC**	**75**	**86.2098**	**0.861**	**1.7755**
	5	**3C**	**75**	**86.1103**	**0.861**	**1.7745**
	6	**χ^2**	10	86.4089	0.863	1.7340
	7	FD	0	86.2525	0.862	1.7245
	8	GI	0	86.2383	0.862	1.7243
	9	RC	56	78.0495	0.775	1.5946
	10	RG	31	37.859	0.364	0.7642
	11	IM	0	26.2297	0.182	0.4442
SVM	1	**GI-SC**	**75**	**87.4325**	**0.874**	**1.8007**
	2	**3C-SC**	**75**	**87.3187**	**0.873**	**1.7986**
	3	**RDFC**	**75**	**87.3045**	**0.873**	**1.7984**
	4	**3C**	**75**	**87.2903**	**0.873**	**1.7983**
	5	**χ^2-SC**	**75**	**86.9918**	**0.87**	**1.7923**
	6	**χ^2**	10	87.1766	0.872	1.7507
	7	GI	0	87.3187	0.873	1.7461
	8	FD	0	87.3187	0.873	1.7461
	9	RC	56	76.0165	0.777	1.5763
	10	RG	31	36.0108	0.339	0.7207
	11	IM	0	26.1871	0.18	0.4418

Tableau 6.10 Résultats avec un corpus de 1000 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**3C**	**100**	**86.7785**	**0.867**	**1.8047**
	2	**3C-SC**	**100**	**86.7643**	**0.867**	**1.8045**
	3	**RDFC**	**100**	**86.7643**	**0.867**	**1.8045**
	4	**χ^2-SC**	**100**	**86.7359**	**0.867**	**1.8042**
	5	**GI-SC**	**100**	**86.7359**	**0.867**	**1.8042**
	6	**χ^2**	26	86.878	0.868	1.7549
	7	GI	0	86.7501	0.867	1.7345
	8	FD	0	86.6648	0.866	1.7326
	9	RC	100	79.9261	0.796	1.6651
	10	RG	31	38.2997	0.368	0.7726
	11	IM	0	26.5852	0.186	0.4518
SVM	1	**χ^2-SC**	**100**	**87.6742**	**0.877**	**1.8236**
	2	**GI-SC**	**100**	**87.6031**	**0.876**	**1.8219**
	3	**3C-SC**	**100**	**87.5604**	**0.876**	**1.8215**
	4	**RDFC**	**100**	**87.532**	**0.875**	**1.8202**
	5	**3C**	**100**	**87.3045**	**0.873**	**1.8159**
	6	**χ^2**	26	87.5036	0.875	1.7682
	7	FD	0	87.333	0.873	1.7463
	8	GI	0	87.205	0.872	1.7440
	9	RC	100	77.8362	0.792	1.6402
	10	RG	31	36.1672	0.341	0.7243
	11	IM	0	26.6562	0.187	0.4535

Tableau 6.11 Résultats avec un corpus de 1250 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**χ^2-SC**	**100**	**87.1624**	**0.871**	**1.8125**
	2	**GI-SC**	**100**	**87.1481**	**0.871**	**1.8124**
	3	**3C-SC**	**100**	**87.1055**	**0.87**	**1.8109**
	4	**RDFC**	**100**	**87.1055**	**0.87**	**1.8109**
	5	χ^2	53	85.5843	0.856	1.7489
	6	GI	5	86.9918	0.869	1.7424
	7	**3C**	**5**	**86.9633**	**0.869**	**1.7421**
	8	FD	3	86.9065	0.868	1.7391
	9	RC	105	81.433	0.811	1.6987
	10	RG	31	38.5129	0.368	0.7748
	11	IM	1	27.7794	0.199	0.4774
SVM	1	**3C-SC**	**100**	**87.6457**	**0.876**	**1.8223**
	2	**RDFC**	**100**	**87.6315**	**0.876**	**1.8222**
	3	**GI-SC**	**100**	**87.6031**	**0.876**	**1.8219**
	4	**χ^2-SC**	**100**	**87.5889**	**0.876**	**1.8218**
	5	χ^2	53	86.4231	0.864	1.7652
	6	GI	5	87.6884	0.877	1.7573
	7	**3C**	**5**	**87.4893**	**0.875**	**1.7533**
	8	FD	3	87.5462	0.875	1.7525
	9	RC	105	78.6039	0.798	1.6574
	10	RG	31	36.0535	0.339	0.7212
	11	IM	1	26.2866	0.218	0.4815

Tableau 6.12 Résultats avec un corpus de 1500 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**χ^2-SC**	**100**	**87.205**	**0.871**	**1.8129**
	2	**3C-SC**	**100**	**87.1624**	**0.871**	**1.8125**
	3	**RDFC**	**100**	**87.0205**	**0.871**	**1.8111**
	4	**GI-SC**	**100**	**87.1197**	**0.87**	**1.8111**
	5	**χ^2**	50	87.404	0.874	1.7830
	6	GI	10	87.2334	0.872	1.7513
	7	**3C**	**12**	**87.1908**	**0.871**	**1.7512**
	8	FD	9	87.0771	0.87	1.7470
	9	RC	105	81.6036	0.813	1.7024
	10	RG	37	41.7117	0.405	0.8479
	11	IM	150	28.4333	0.208	0.5972
SVM	1	**χ^2-SC**	**100**	**87.9016**	**0.879**	**1.8279**
	2	**RDFC**	**100**	**87.7168**	**0.877**	**1.8240**
	3	**GI-SC**	**100**	**87.7168**	**0.877**	**1.8240**
	4	**3C-SC**	**100**	**87.7026**	**0.877**	**1.8239**
	5	**χ^2**	50	87.9016	0.879	1.7929
	6	**3C**	**12**	**87.6884**	**0.877**	**1.7622**
	7	FD	9	87.7168	0.877	1.7604
	8	GI	10	87.6457	0.876	1.7594
	9	RC	105	78.9309	0.801	1.6637
	10	RG	37	38.6125	0.369	0.7809
	11	IM	150	26.4714	0.222	0.5916

Tableau 6.13 Résultats avec un corpus de 1750 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**χ^2-SC**	**350**	**87.3472**	**0.873**	**1.9912**
	2	**χ^2-SC**	**300**	**87.3756**	**0.873**	**1.9565**
	3	**GI-SC**	**300**	**87.333**	**0.873**	**1.9561**
	4	**3C-SC**	**175**	**87.4183**	**0.873**	**1.8695**
	5	**RDFC**	**175**	**87.3898**	**0.873**	**1.8692**
	6	**χ^2**	56	87.6315	0.876	1.7914
	7	**3C**	**56**	**87.532**	**0.875**	**1.7894**
	8	GI	26	87.2334	0.872	1.7625
	9	FD	24	87.2619	0.872	1.7614
	10	RC	129	82.1439	0.818	1.7296
	11	RG	82	46.3748	0.46	0.9810
	12	IM	361	29.4569	0.225	0.7720
SVM	1	**χ^2-CS**	**350**	**87.9585**	**0.879**	**2.0033**
	2	**χ^2-CS**	**300**	**87.9727**	**0.88**	**1.9695**
	3	**GI-SC**	**300**	**87.6599**	**0.876**	**1.9623**
	4	**3C-SC**	**175**	**87.7452**	**0.877**	**1.8768**
	5	**RDFC**	**175**	**87.6457**	**0.876**	**1.8748**
	6	**χ^2**	56	87.9727	0.88	1.7988
	7	**3C**	**56**	**87.6599**	**0.876**	**1.7917**
	8	FD	24	87.4893	0.875	1.7666
	9	GI	26	86.6884	0.877	1.7620
	10	RC	129	79.4569	0.805	1.6897
	11	RG	82	41.8396	0.41	0.8857
	12	IM	361	26.9406	0.231	0.7528

Tableau 6.14 Résultats avec un corpus de 2000 attributs

Algorithme	Position	Méthode appliquée	# d'attributs écrits en caractères latins et conservés	Exactitude (%)	F-1	Eval_Multi
NBM	1	**χ^2-SC**	**300**	**87.6599**	**0.876**	**1.9623**
	2	**GI-SC**	**300**	**87.5604**	**0.875**	**1.9603**
	3	**3C-SC**	**200**	**87.7168**	**0.877**	**1.8940**
	4	**GI-SC**	**200**	**87.6884**	**0.876**	**1.8927**
	5	**RDFC**	**200**	**87.6742**	**0.876**	**1.8926**
	6	**χ^2-SC**	**200**	**87.6742**	**0.876**	**1.8926**
	7	**χ^2**	172	87.7737	0.877	1.8750
	8	RC	221	84.3759	0.842	1.8403
	9	FD	62	87.6742	0.876	1.7960
	10	**3C**	**58**	**87.7026**	**0.876**	**1.7935**
	11	GI	58	87.6884	0.876	1.7934
	12	RG	413	46.8439	0.466	1.2232
	13	IM	716	30.9923	0.247	1.0576
SVM	1	**χ^2-SC**	**200**	**87.9016**	**0.879**	**1.8978**
	2	**GI-SC**	**200**	**87.8732**	**0.878**	**1.8965**
	3	**3C-SC**	**200**	**87.8163**	**0.878**	**1.8960**
	4	**RDFC**	**200**	**87.6031**	**0.878**	**1.8938**
	5	**χ^2**	172	87.779	0.878	1.8760
	6	RC	221	82.9542	0.832	1.8160
	7	FD	62	87.9443	0.879	1.8017
	8	GI	58	87.9585	0.88	1.8001
	9	**3C**	**58**	**87.8732**	**0.879**	**1.7982**
	10	RG	413	42.2235	0.415	1.1260
	11	IM	716	28.021	0.25	1.0309

Tableau 6.15 Résultats avec un corpus de 2500 attributs

Annexe des algorithmes des solutions proposées

Fonction 3C ( )

{

$j = 1$;
Soit l une liste où les attributs pondérés seront stockés;
$l = \emptyset$;

Tant que $j \leq le\ nombre\ d'attributs$
{
Soit a_j un attribut de l'espace d'apprentissage ;
Soit $3C(a_j) = 0$;

Pour chaque catégorie c_i
{
A_{ij} = Le nombre de documents $\in c_i$ contenant l'attribut a_j ;
C_{ij} = Le nombre de documents $\in c_i$ ne contenant pas l'attribut a_j ;

$3C_i(a_j) = \frac{A_{ij}}{A_{ij}+C_{ij}}$;

$3C(a_j) = 3C_i(a_j) + 3C(a_j)$;
Ajouter a_j avec son score à l;
}
$j = j + 1$;
}

Trier la liste l en ordre décroissant de scores ;
Sélectionner <u>*séquentiellement*</u> les premiers n attributs ;
}

Algorithme 6.1 *L'algorithme de la méthode « 3C »*

Fonction 3C-SC ( )

{

$j = 1$;
Soit l une liste où les attributs pondérés seront stockés;
$l = \emptyset$;

Tant que $j \leq le\ nombre\ d'attributs$
{
Soit a_j un attribut de l'espace d'apprentissage ;
Soit $3C(a_j) = 0$;

Pour chaque catégorie c_i
{
A_{ij} = Le nombre de documents $\in c_i$ contenant l'attribut a_j ;
C_{ij} = Le nombre de documents $\in c_i$ ne contenant pas l'attribut a_j ;

$3C_i(a_j) = \frac{A_{ij}}{A_{ij}+C_{ij}}$;

$3C(a_j) = 3C_i(a_j) + 3C(a_j)$;

Ajouter a_j avec son score à l;
}
$j = j + 1$;
}

Trier la liste l en ordre décroissant de scores ;
Sélectionner les premiers n attributs de l selon la façon suivante :

- Sélectionner de l les premiers na attributs écrits en caractères arabes;
- Sélectionner de l les premiers nl attributs écrits en caractères latins tel que $n = na + nl$;

}

Algorithme 6.2 *L'algorithme de la méthode « 3C-SC »*

Fonction RFDC ( )
{
$j = 1$;
Soit la la liste d'attributs écrits en caractères arabes ;
Soit ll la liste d'attributs écrits en caractères latins ;
$la = \emptyset;\ ll = \emptyset;$
Tant que $j \leq le\ nombre\ d'attributs$
{
Soit a_j un attribut de l'espace d'apprentissage ;
Si (a_j est écrit en caractères arabes)
{
RFDC $(a_j) = 3C(a_j)$;
Ajouter a_j avec son score à la ;
}
Sinon
{
Soit RFDC $(a_j) = 0$;
$A_j = 0;$
$C_j = 0;$
$ICF_j = 0;$
Pour chaque catégorie c_i
{
A_{ij} = Le nombre de documents $\in c_i$ contenant l'attribut a_j ;
C_{ij} = Le nombre de documents $\in c_i$ ne contenant pas l'attribut a_j ;
$A_j = A_j + A_{ij}$;
$C_j = C_j + C_{ij}$;
$ICF_{ij} = \frac{A_{ij}}{A_{ij}+C_{ij}} . \left(\log_2 \frac{A_{ij}}{A_{ij}+C_{ij}}\right)$, $\forall A_{ij} = 0 \rightarrow ICF_{ij} = 0$;
$ICF_j = ICF_j + ICF_{ij}$;
}
RFDC $_i(a_j) = \frac{A_j}{C_j} + ICF_j$;
Ajouter a_j avec son score à ll ;
}
$j = j + 1$;
}
Trier la et ll en ordre décroissant de scores ;
Sélectionner les premiers $n = na + nl$ attributs des deux listes tel que :
- na est le nombre des premiers attributs dans la;
- nl est le nombre des premiers attributs dans ll;

}

Algorithme 6.3 *L'algorithme de la méthode « RFDC »*

Extrait des mots vides

نحن	لن	اللاتي	هنا	أولاء	من
أنتم	إذن	الألاء	عل	هؤلاء	إلى
أنتن	لو	اللائي	مع	أولائك	عن
هم	لم	ذو	هناك	أبد	على
هن	لا	كلا	هنالك	عند	في
عسى	إن	دون	حين	آناء	رب
أجل	لولا	كذا	كم	إزاء	خلا
ألا	هلا	إلا	كيف	آن	عدا
بلى	ألا	غير	أي	أين	حاشا
بله	قد	بيد	من	ريثما	كي
سواء	ثم	سوى	ما	متى	حتى
كلا	أو	أيها	ماذا	مذ	ذا
كأي	أم	يا	أينما	منذ	هذا
كل	بل	أيا	حيثما	لما	ذي
كلتا	هل	وا	كيفما	إذ	هذه
لات	أي	أيتها	مهما	إذا	ذاك
لوما	أما	إن	الذي	أنى	ذلك
لكن	أنا	أن	التي	أيان	تلك
نعم	إيا	كأن	اللذان	بعد	هذان
ها	أنت	لكن	اللذين	قبل	هذين
بعض	أنت	ليت	اللتان	بين	هاتان
لدن	هو	لعل	اللتين	بينما	هاته
ثم	هي	أما	اللذين	عوض	هاتين
ثمة	أنتما	إما	اللواتي	قط	أولى

Transcription des caractères arabes

Les emphatiques et la constrictive vélaire ***hâ'*** sont en caractère gras. Le soulignement distingue, lorsqu'il y a lieu, la constrictive de l'occlusive correspondante, ou une consonne d'une consonne "voisine" ; les voyelles longues portent un accent circonflexe. On a :

- Voyelles brèves : a, u, i
- Voyelles longues : â, û, î.
- Consonnes (par ordre alphabétique) : *hamza* = ' ; *bâ'* = b ; *tâ'* = t ; *t̲â'* = t̲ ; *jîm* = j ; ***h***â' = **h** ; *xâ'* = x ; *dâl* = d ; *d̲â* =d̲ ; *râ'* = r ; *zây* = z ; *sîn* = s ; *s̲în* = s̲ ; ***s***âd = **s** ; ***d***âd = **d** ; ***t***â' = **t** ; ***d̲***âl' = **d̲** ; *'ayn* = ' ; *g̲ayn* = g̲ ; *fa'* = f ; *qâf* = q ; *kâf* = k ; *lâm* = l ; *mîm* = m ; *nûn* = n ; *hâ'* = h ; *wâw* = w ; *yâ'* = y.
- Morphogramme : *tâ' marbû**t**a* = &.
- Représentation des mots graphiques : les symboles ≤ et ≥ encadrent les notations en translittération graphique, qui "imitent" la représentation graphémique de l'arabe : les *majuscules* transcrivent les graphèmes inclus dans le corps du mot, et les *minuscules,* les graphèmes diacritiques secondaires (également appelés "signes de vocalisation).

Un exemple d’une analyse morphologique d’un document

Le texte original (après l’élimination des mots vides):

جاء بيان اصدرته الحكومة العراقية الثلاثاء العراق قرر شراء طائرة لنقل الركاب شركة بوينج الامريكية لاستخدامها شركة الخطوط الجوية العراقية المملوكة للدولة وتقول وكالة رويترز للانباء العراق قرر شراء ست طائرات شركة بومباردير الكندية لاستخدامها الرحلات القصيرة يعلن قيمة العقود طراز الطائرات قرر العراق شرائها.

Les lemmes :

جاء بيان اصدر حكومة عراقي الثلاثاء العراق قرر شراء طائرة نقل راكب شركة بوينج امريكي استخدام شرك خط جوي عراقي مملوك دولة قول وكالة رويترز أنباء العراق قرر شراء ست طائرة شركة بومباردير كندي استخدام رحلة قصير علن قيمة عقد طراز طائرة قرر العراق شراء.

Les racines :

جيء بين صدر حكم العراق الثلاثاء العراق قرر شري طير نقل ركب شرك بوينج الامريكية خدم شرك خطط جوو العراق ملك دول قول وكل رويترز نبء العراق قرر شري ستة طير شرك بومباردير الكندية خدم رحل قصر علن قوم عقد طرز طير قرر العراق شري.

Mot original	**Lemme**	**Racine**
جاء	جاء	جيء
بيان	بيان	بين
اصدرته	اصدر	صدر
الحكومة	حكومة	حكم
العراقية	عراقي	العراق
قرر	قرر	قرر
شراء	شراء	شري
طائرة	طائرة	طير
لنقل	نقل	نقل
الركاب	راكب	ركب
شركة	شرك	شرك
بوينج	بوينج	بوينج
الامريكية	امريكي	الامريكية
لاستخدامها	استخدام	خدم
…		