

HIROSHI AKUTO

LUDOVIC LEBART

**Le repas idéal : analyse de réponses libres en  
trois langues : anglais, français, japonais**

*Les cahiers de l'analyse des données*, tome 17, n° 3 (1992),  
p. 327-352

[http://www.numdam.org/item?id=CAD\\_1992\\_\\_17\\_3\\_327\\_0](http://www.numdam.org/item?id=CAD_1992__17_3_327_0)

© Les cahiers de l'analyse des données, Dunod, 1992, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

# LE REPAS IDÉAL: ANALYSE DE RÉPONSES LIBRES EN TROIS LANGUES: ANGLAIS, FRANÇAIS, JAPONAIS

## [REPAS IDÉAL]

*Hiroshi AKUTO\**  
*Ludovic LEBART\*\**

### 0 Des données aux analyses

Une enquête internationale a eu lieu en 1989-1990, à l'initiative de l'*Institute of Research on Urban Life* (Institut de Recherche Japonais dépendant de la *Tokyo Gas Company Ltd*) sur les comportements, habitudes et préférences alimentaires de trois grandes métropoles (et mégapoles) : Paris, New-York et Tokyo.

L'enquête a donné lieu à environ un millier d'entrevues (entretiens face à face) dans chacune de ces trois villes.

Le questionnaire comprenait de nombreuses questions fermées décrivant de façon détaillée le rythme, les horaires, les lieux et circonstances des repas ou collations. La composition des repas habituels était relevée, ainsi que les préférences alimentaires dans la vie quotidienne et lors de circonstances exceptionnelles. À cela, s'ajoutaient deux questions ouvertes :

Quels sont les plats que vous aimez et mangez fréquemment ? (avec une relance : Y-a-t-il d'autre plats que vous aimiez et mangiez fréquemment ?)

Pouvez-vous donner des exemples de ce que vous considérez comme un "repas idéal" ? (avec la relance : Y-a-t-il d'autres choses qui feraient partie de votre "repas idéal" ?)

Dans les premières analyses qui sont présentées ici, les réponses à ces deux questions sont regroupées, et considérées comme une seule réponse. Ce regroupement permet d'avoir des réponses individuelles relativement riches, au prix d'une certaine perte d'homogénéité.

---

(\*) Professeur à l'Université de Tokyo, Département de Psychologie Sociale.

(\*\*) Directeur de Recherche au CNRS, Ecole Nationale Supérieure des Télécommunications.

Les questions ouvertes peuvent être utilisées à de multiples fins lors des enquêtes. Trois usages prédominent :

acquérir une information étendue en un temps très court

explicitier les réponses à des questions fermées. C'est le cas de la question "pourquoi ?", qui réserve bien des surprises lors de l'interprétation des résultats... En effet, on suppose généralement que toutes les personnes interrogées comprennent de la même façon les questions qui leur sont posées. L'analyse des réponses à la question ouverte "pourquoi ?" remet souvent en cause cet *a priori*.

explorer un domaine peu connu, où les attitudes des sujets sont imprévisibles, en espérant découvrir en même temps que mesurer.

C'est plutôt à ce dernier type d'usage qu'on destine les réponses aux questions ouvertes précitées.

Les analyses de type lexical sont familières aux lecteurs des *Cahiers*, c'est pourquoi nous serons brefs sur la méthodologie générale de l'application qui va suivre.

L'unité statistique choisie sera la forme graphique, séquence de caractères séparés par des caractères délimiteurs (blancs, virgules, points, parenthèses, etc.). À cette fin, les réponses en Japonais ont été romanisées.

Les trois langues seront traitées séparément. Mais les profils lexicaux de chacune des langues peuvent caractériser des catégories communes, comme le sexe et l'âge, ou approximativement communes comme le niveau d'instruction, le niveau de vie, la profession.

Bien que repérées au départ dans des espaces différents, dans lesquels les points initiaux représentent des individus anonymes, les points qui représentent ces catégories pourront être caractérisés de façon plus intrinsèque par leurs distances respectives, et par les configurations géométriques qui résument ces distances. À partir d'individus inconnus et de mots inconnus, (au moins considérés comme tels dans une première phase), on peut espérer observer et comparer des configurations de catégories connues.

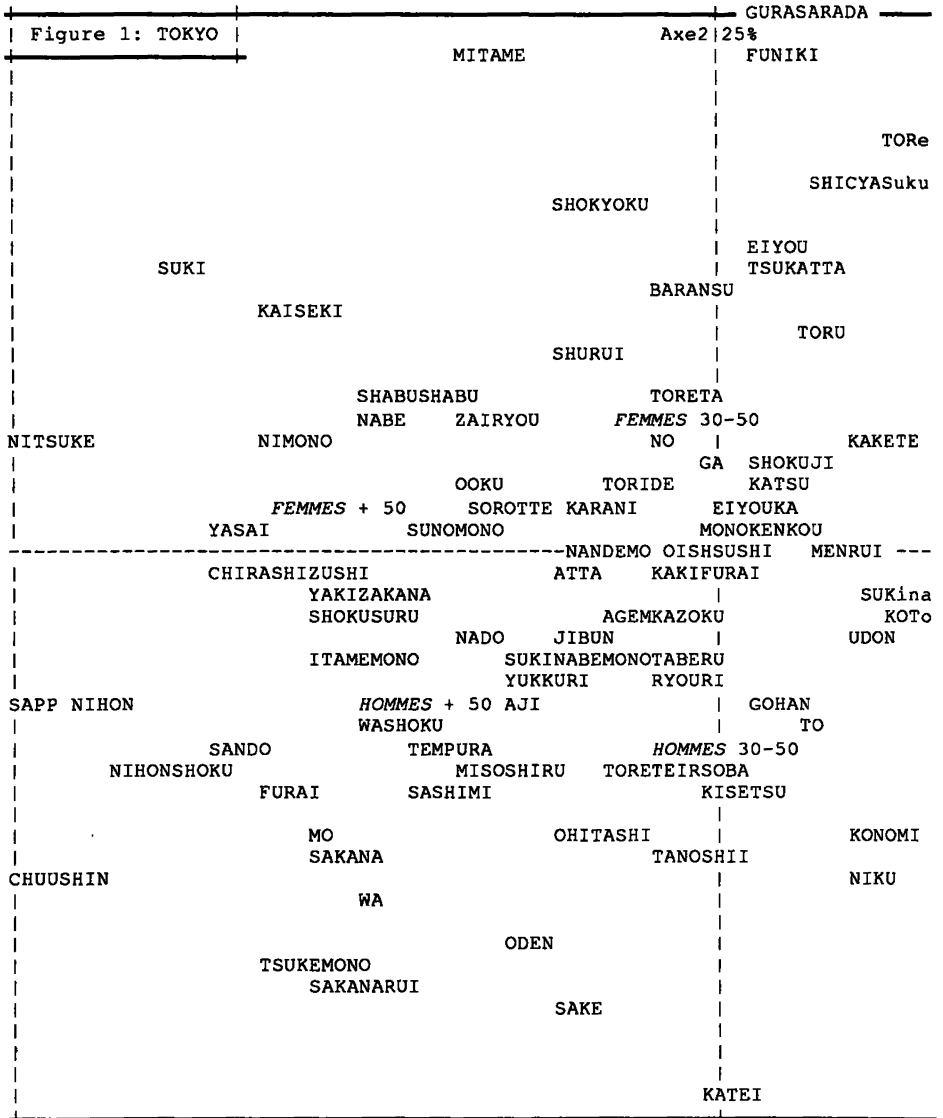
On peut donc procéder à des comparaisons internationales non dépourvues de finesse, comme cette contribution va tenter de le montrer, sans qu'il soit nécessaire de procéder à la moindre traduction pendant la phase de traitement statistique.

Les trois premières parties de cet article seront consacrées à la description des données et des premières analyses pour chacun des trois pays. La quatrième partie est un essai de discussion des résultats obtenus, discussion suscitée notamment par plusieurs remarques de J.-P. Benzécri.

num	Mot	Employé	frq	46	KISETSU	14	93	SHOKUSURU	7
1	AGEMONO	14	47	KONOMI	7	94	SHURUI	10	
2	AJI	21	48	KOTO	128	95	SOBA	10	
3	ANKA	7	49	MENRUI	9	96	SOROTTE	24	
4	ARI	15	50	MISOSHIRU	26	97	SOZAI	20	
5	ARU	40	51	MITAME	14	98	SUKI	9	
6	ATTA	11	52	MO	9	99	SUKINA	47	
7	BARANSU	152	53	MONO	209	100	SUKIYAKI	50	
8	CHAHAN	7	54	NABE	14	101	SUNOMONO	9	
9	CHIRASHIZUSHI	7	55	NABEMONO	54	102	SUPAGETTEI	42	
10	CHUUKA	41	56	NADO	7	103	SURU	24	
11	CHUUSHIN	9	57	NAI	11	104	SUSHI	65	
12	DE	127	58	NAKU	7	105	SUTEKI	13	
13	DEKIRU	8	59	NANDEMO	15	106	TABERAREREBABA	8	
14	DK	14	60	NI	97	107	TABERARERU	26	
15	EIYOU	90	61	NIHON	44	108	TABERU	127	
16	EIYOUKA	8	62	NIHONSHOKU	21	109	TABETAI	25	
17	FUNKI	10	63	NIHONSOBA	7	110	TANOSHII	18	
18	FURAI	14	64	NIKU	36	111	TANOSHIKU	37	
19	FURANSU	10	65	NIKURUI	10	112	TEMPURA	38	
20	GA	106	66	NIMONO	96	113	TENKABUTSU	8	
21	GOHAN	11	67	NITSUKE	9	114	TO	92	
22	GURATAN	20	68	NIZAKANA	20	115	TOKI	12	
23	GYOUZA	15	69	NO	300	116	TOMONI	12	
24	HAMBAGA	7	70	O	223	117	TONKATSU	22	
25	HAMBAGU	42	71	ODEN	15	118	TORE	15	
26	IROIRONO	7	72	OHITASHI	9	119	TORETA	73	
27	ISSHONI	14	73	OISHII	95	120	TORETEIRU	8	
28	ITAMEMONO	14	74	OISHIKU	75	121	TORI	8	
29	JIBUN	27	75	OUI	7	122	TORIIIRETA	7	
30	JIKAN	12	76	OOKU	16	123	TORU	23	
31	KAISEKI	8	77	PIZA	11	124	TSUKATTA	8	
32	KAKETE	8	78	RAISU	15	125	TSUKEMONO	14	
33	KAKIFURAI	7	79	RAMEN	27	126	UDON	15	
34	KARAGE	11	80	RYOURI	258	127	WA	17	
35	KARADA	15	81	SAKANA	63	128	WASHOKU	87	
36	KARE	79	82	SAKANARUI	14	129	YAKINIKU	62	
37	KARORI	11	83	SAKE	13	130	YAKITORI	8	
38	KATEI	14	84	SANDO	12	131	YAKIZAKANA	57	
39	KATSU	7	85	SAPPARISHITA	8	132	YASAI	95	
40	KATSUDON	7	86	SARADA	20	133	YASAIITAME	29	
41	KAZOKU	64	87	SASHIMI	94	134	YASUKU	10	
42	KENKOU	28	88	SHABUSHABU	7	135	YOI	62	
43	KENKOUTEKI	8	89	SHICHUU	21	136	YOKU	42	
44	KENKOUTEKINA	7	90	SHINSENNNA	11	137	YUKKURI	17	
45	KI	11	91	SHOKUHIN	8	138	YUJIN	12	
			92	SHOKUJI	184	139	ZAIRYOU	16	

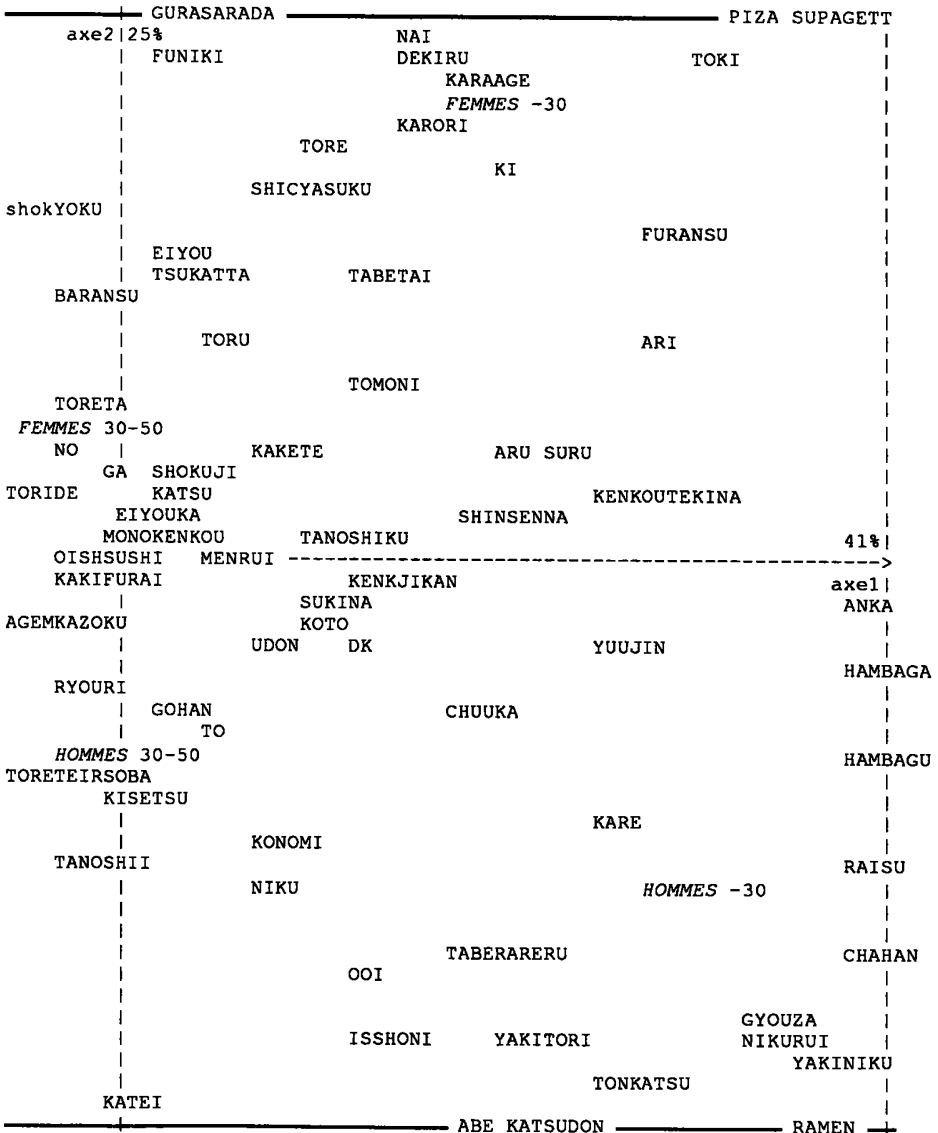
Tableau 1 : TOKYO : formes graphiques par ordre alphabétique ; seuil = 6

Deux variables élémentaires apparaissent jouer un rôle privilégié dans les trois pays concernés : l'âge et le sexe. Le choix de ces variables sera évoqué de nouveau dans la quatrième partie de cet article. Mais c'est finalement une variable composite comprenant six catégories (trois catégories d'âge, croisées avec le sexe) qui nous servira à intégrer les réponses reçues dans les trois pays.



**1 Sous-échantillon de Tokyo**

Le texte total des 1008 réponses contient 6219 occurrences de 832 formes distinctes (dont le nombre représente 13,4% du texte). Le tableau 1 montre les 139 formes apparaissant au moins 7 fois dans l'échantillon des 1008 réponses aux deux questions ouvertes. Ces 139 formes donnent lieu, en tout, à 4975



occurrences, soit 80% du texte total. On note la présence de mots outil, (conjonction, préposition comme TO, NO, NI, DE ), de mots étrangers adoptés par la langue japonaise, puis romanisés pour les besoins de cette analyse, ce qui donne les transcriptions suivantes: BARANSU (Balanced: équilibré), GURATAN (gratin), HAMBAGU, HAMBAGA (deux variétés du mot

Hamburger), KARE (Curry), RAISU (riz), SARADA (Salade), PIZA, SUPAGETTEI, SUTEKI.

Apparaissent également quelques classiques de la cuisine japonaise, maintenant bien connus des occidentaux comme les SASHIMI, SUSHI, SUKIYAKI, YAKITORI, TEMPURA. D'autres formes seront commentées plus loin.

La romanisation introduit une confusion que permettait de trancher les Kanjis chargés de sens ou même les accents: SAKE, par exemple, désigne aussi bien le vin de riz (connu sous le même nom en Français) que le saumon. Ce type de mutilation de l'information de base n'enlève que peu de chose à la richesse des profils lexicaux, comme on va en juger.

La figure 1 présente le plan (1, 2) issu de l'analyse du tableau de contingence à 139 lignes et 6 colonnes obtenu en groupant les 1008 profils en 6 classes correspondant à la variable âge-sexe évoquée plus haut:

Moins de 30 ans	Hommes	De 30 à 50 ans	Hommes	Plus de 50 ans	Hommes
Moins de 30 ans	Femmes	De 30 à 50 ans	Femmes	Plus de 50 ans	Femmes

Ce regroupement *a priori* privilégie les variables choisies: mais c'est une analyse sans *a priori* des 1038 profils qui nous a incité à procéder ainsi.

Première observation : les profils lexicaux permettent de reconstituer sans interversion les positions relatives des classes : l'âge augmente de la droite vers la gauche, les hommes occupent le bas du graphique, les femmes la partie supérieure. Le premier axe (41% de l'inertie) est donc un axe d'âge, le second (25% de l'inertie), un axe qui sépare les sexes.

Du côté des jeunes, et surtout des garçons, les hamburgers, de la viande grillée, (YAKI = grillé, NIKU = viande, YAKITORI = poulet grillé), du riz, des nouilles (RAMEN). Pour les jeunes filles ou femmes, on peut noter PIZA, SUPPAGETTEI, FURANSU (français). D'une manière générale, les jeunes sont surtout ceux qui s'intéressent à la cuisine étrangère, les filles étant peut-être plus européennes que les garçons.

Du côté des classes les plus âgées, on trouve bien sûr NIHON (japonais), et les diverses sortes ou préparations de poissons (SAKANA ou ZAKANA), avec mention de la KAISEKI, cuisine Kyotoïte très raffinée d'inspiration Zen.

Si les distances représentées dans le plan (1, 2) sont une bonne représentation des distances dans l'espace initial à cinq dimensions (il y a six catégories), on constate de plus une convergence progressive des points relatifs aux deux sexes pour une même classe d'âge lorsque l'âge augmente, ce qui donne une forme trapézoïdale à la configuration. Nous reviendrons sur cette constatation plus loin. Les explications ne manquent pas à cette observation : traditionalisme, vie commune et "fusionnelle", et donc homogénéité des

	LIBELLE DE LA FORME GRAPHIQUE	POURCENTAGE INTERNE GLOBAL	FREQUENCE INTERNE GLOBALE	V.TEST	PROBA		
<i>moins de 30 ans - HOMMES</i>							
1	YAKINIKU	3.95	1.25	26.	62.	5.509	0.000
2	KARE	3.95	1.59	26.	79.	4.427	0.000
3	HAMBAGU	2.58	0.84	17.	42.	4.268	0.000
4	CHAHAN	0.91	0.14	6.	7.	3.991	0.000
5	RAMEN	1.82	0.54	12.	27.	3.812	0.000
6	RAISU	1.21	0.30	8.	15.	3.483	0.000
<i>de 30 à 50 ans - HOMMES</i>							
1	SAKE	0.70	0.26	7.	13.	2.455	0.007
2	KAZOKU	2.10	1.29	21.	64.	2.295	0.011
3	WA	0.80	0.34	8.	17.	2.290	0.011
4	KISETSU	0.70	0.28	7.	14.	2.267	0.012
5	SASHIMI	2.71	1.89	27.	94.	1.927	0.027
6	SUTEKI	0.60	0.26	6.	13.	1.877	0.030
<i>plus de 50 ans - HOMMES</i>							
1	SASHIMI	4.00	1.89	23.	94.	3.421	0.000
2	SAKANA	2.78	1.27	16.	63.	2.938	0.002
3	CHUUSHIN	0.87	0.18	5.	9.	2.926	0.002
4	NIHON	2.09	0.88	12.	44.	2.720	0.003
5	NABEMONO	1.91	1.09	11.	54.	1.731	0.042
6	YASAI	2.96	1.91	17.	95.	1.720	0.043
<i>moins de 30 ans - FEMMES</i>							
1	SUPAGETTEI	3.57	0.84	24.	42.	6.536	0.000
2	PIZA	1.04	0.22	7.	11.	3.595	0.000
3	SARADA	1.34	0.40	9.	20.	3.239	0.001
4	GURATAN	1.04	0.40	7.	20.	2.237	0.013
5	EIYOU	2.97	1.81	20.	90.	2.160	0.015
6	TOKI	0.74	0.24	5.	12.	2.155	0.016
<i>30 à 50 ans - FEMMES</i>							
1	BARANSU	4.82	3.06	56.	152.	3.710	0.000
2	NABE	0.69	0.28	8.	14.	2.476	0.007
3	O	5.85	4.48	68.	223.	2.430	0.008
4	NAKU	0.43	0.14	5.	7.	2.347	0.009
5	GURATAN	0.77	0.40	9.	20.	1.927	0.027
6	TORE	0.60	0.30	7.	15.	1.746	0.040
<i>plus de 50 ans - HOMMES</i>							
1	YASAI	4.19	1.91	38.	95.	4.910	0.000
2	NIMONO	3.64	1.93	33.	96.	3.712	0.000
3	WASHOKU	3.09	1.75	28.	87.	3.056	0.001
4	NITSUKE	0.66	0.18	6.	9.	2.905	0.002
5	NIHON	1.76	0.88	16.	44.	2.719	0.003
6	JIBUN	1.21	0.54	11.	27.	2.559	0.005

**Tableau 2: Choix des formes caractéristiques (TOKYO)**

personnes plus âgées, ouverture à la fois culturelle (goûts éclectiques) et matérielle (non cohabitation ou cohabitation partielle, fréquentation des restaurants, invitations) des plus jeunes, qui favoriserait une différenciation des goûts et des habitudes alimentaires.

Reste cependant à vérifier la vérité du fait statistique: la réalité de ce trapèze qui n'est peut-être que la projection déformée d'une réalité autre.

Le tableau 2 complète cette figure en faisant apparaître les formes caractéristiques de chaque catégorie. Pour lire ce tableau, prenons par exemple la



*moins de 30 ans - HOMMES*

- 1 YAKINIKU, KARE RAISU, RAMEN
- 1 OISHII TO OMOTTE TABERARERU KOTO
- 2 RAMEN, YAKINIKU, KARE
- 2 OISHIKUTE RYOU NO ARU KOTO
- 3 KARE RAISU, HAMBAGU, SUPAGETTEI
- 3 NIKURUI, ABURAPPOI RYOU RI, RAMEN NO YOUNA MONO

*de 30 à 50 ans - HOMMES*

- 1 CHUUKA RYOU RI, KATEI RYOU RI, YAKIZAKANA
- 1 SUKINA SAKE TO YASAI TO SAKANA O TORIIRETA MENJU DE
- 2 SASHIMI, CHUUKA RYOU RI
- 2 DEKIREBA KAZOKU DE SHOKUJI SHITAI
- 3 YAKINIKU, NABEMONO, KISETSU NO SHUN O TORIIRETA SHOKUJI,
- 3 YASUI ZAIRYOU GA HOUFU DE KARADA NI YOI SAKE

*plus de 50 ans - HOMMES*

- 1 NABEMONO, MAGURO NO SASHIMI
- 1 YASAI TO SAKANA GA ISSHONI NATTA RYOU RI
- 2 WASHOKU
- 2 SAKANA, YASAI, TOUFU, NIMONO, JUNNIHONTEKINA MONO
- 3 NABEMONO, TONKATSU, YAKIZAKANA
- 3 YASAI SAKANA CHUUSHIN NO WASHOKU

*moins de 30 ans - FEMMES*

- 1 KARE, GURATAN, PIZA, UNAGI, SUPAGETTEI
- 1 EIYOU NO BARANSU TORETA SHOKUJI
- 2 GURATAN, PIZA, SUPAGETTEI, CHIRASHI, KARE, HAMBAGA, SARADA
- 2 OISHIKU EIYOU GA ARI MAINICHI HENKA NI TONDA KONDATE NO SHOKUJI
- 3 SUPAGETTEI
- 3 BARANSU TORETA SHOKUJI

*de 30 à 50 ans - FEMMES*

- 1 YASAI NO NIMONO
- 1 KAZOKU DE TSUKURITATE NO MONO O YUKKURI KATARINAGARA TABERU SHOKUJI
- 2 GOMOKUGOHAN
- 2 EIYOU NO BARANSU TORETA MONO
- 3 SUSHI
- 3 MITAME GA KIREI DE EIYOU NO BARANSU NO TORETA SHOKUJI

*plus de 50 ans - FEMMES*

- 1 YASAI NO NIMONO
- 1 WASHOKU
- 2 YASAI NO NIMONO, SASHIMI, NIZAKANA, YAKIZAKANA
- 2 JIBUN NI ATTA SUKINA MONO O TABERU KOTO
- 3 YASAI NO NIMONO, KORUFOBIFU, SUTEKI
- 3 JIBUN DE RYOURISHITA MONO O OISHIKU AJIWAWU KOTO GA DEKIREBA RISOU DE ARU

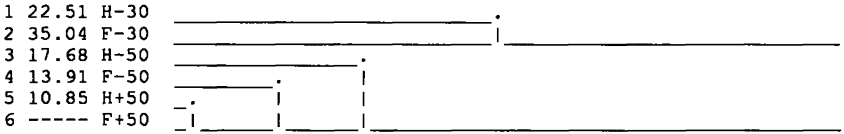
**Tableau 3 : individus ou réponses caractéristiques (TOKYO)**

forme YAKINIKU, qui apparaît caractéristique des hommes de moins de 30 ans. On lit sur ce tableau que le pourcentage des occurrences de cette forme chez les hommes de moins de 30 ans (3,95%) est plus de trois fois supérieur à son pourcentage global (1,25%). L'avant-dernière colonne donne à cet écart une valeur-test de 5,51. (La valeur-test convertit une probabilité critique en variable normale standardisée: la valeur 1,96 correspond ainsi au seuil 0,025, la valeur 5,51 correspond à une probabilité de l'ordre du millionième). La dernière colonne fournit d'ailleurs les premiers chiffres de ce seuil.

On observe que certaines formes, comme GURATAN, caractérisent les femmes des deux premières classes d'âge. En revanche, une forme comme

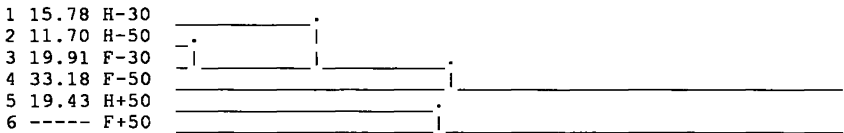
**TOKYO**

RANG IND. IDEN SOMME DES INDICES: 0.25917 MIN = 10.85% / MAX = 35.04%



**PARIS**

RANG IND. IDEN SOMME DES INDICES: 0.10370 MIN = 11.70% / MAX = 33.18%



**NEW YORK**

RANG IND. IDEN SOMME DES INDICES: 0.13618 MIN = 13.64% / MAX = 27.34%

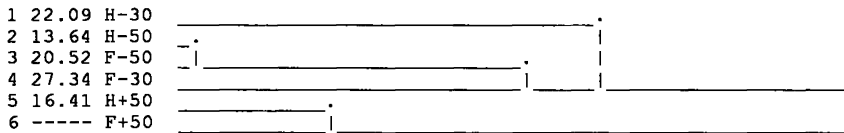
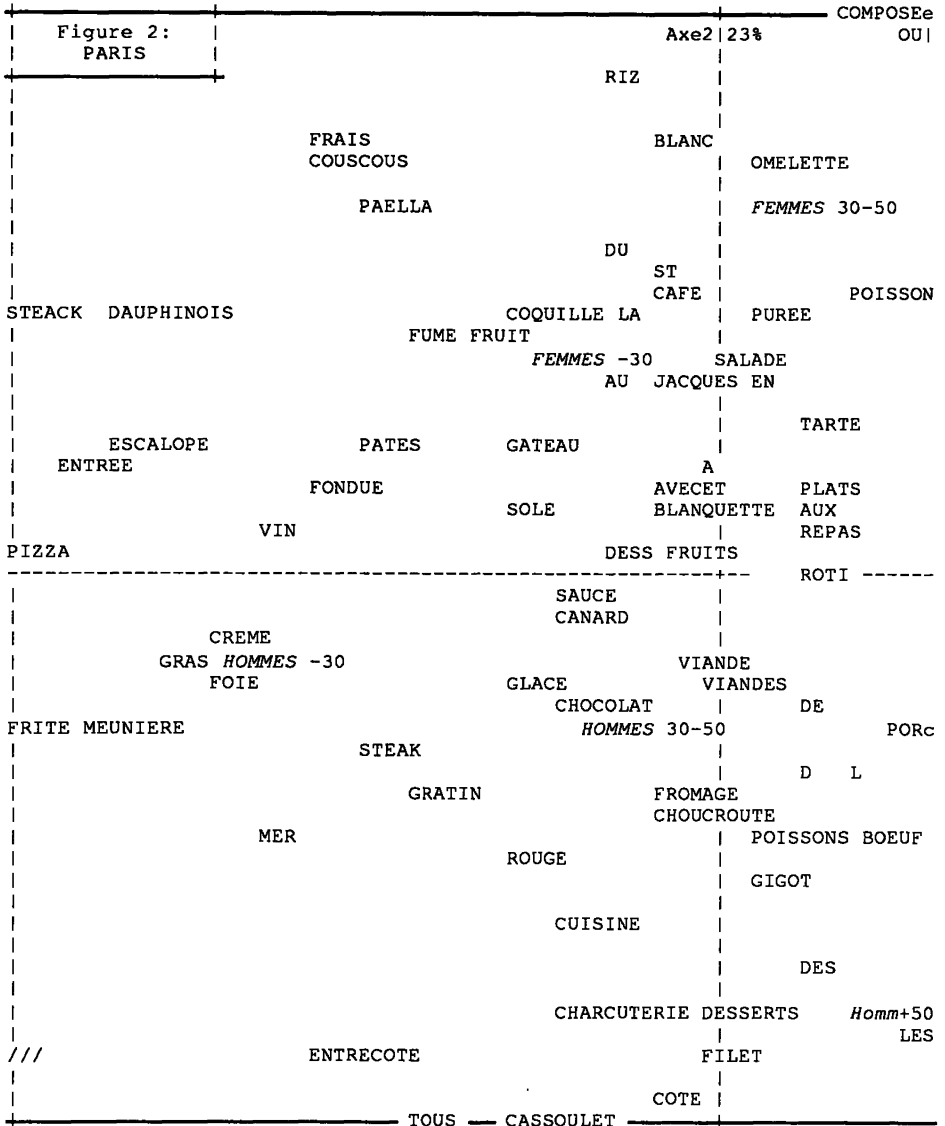


Figure 4: Comparaison des dendrogrammes: TOKYO, PARIS, NEW-YORK  
N.B. La figure ci-dessus sera commentée au §4.1

YASAI (légumes) caractérise les personnes âgées des deux sexes, tout en étant beaucoup plus caractéristique des femmes.

Le tableau 3 donne les réponses caractéristiques de chacune des catégories étudiées. Rappelons qu'il s'agit de réponses originales choisies comme étant les plus proches du centre de cette classe, (au sens de la distance du Chi-deux entre le profil lexical de la réponse et le profil lexical moyen de la classe). Un autre mode de calcul consiste à caractériser une réponse par la valeur-test moyenne des formes qu'elle contient, mais ce critère favorise trop les réponses lapidaires, et n'a pas été utilisé ici.

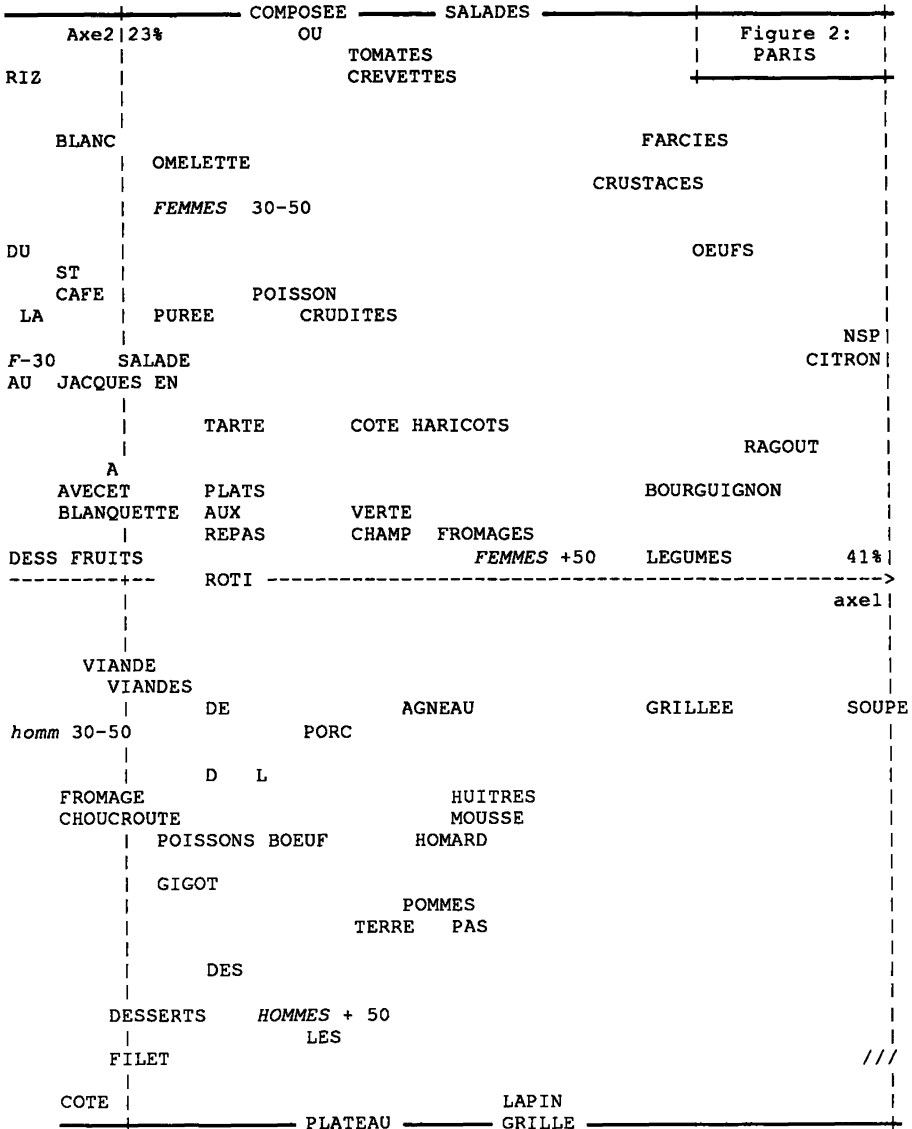
Les réponses caractéristiques sont utiles car elles plongent les formes dans leur contexte réel, et rappellent la complexité de l'information de base. Ainsi, la forme SUPAGETTEI, très caractéristique des jeunes femmes, se rencontre parfois chez leurs homologues masculins: si la réponse 3 est typique des garçons, c'est qu'elle est lestée par ailleurs de formes très "masculines".



**2 Sous-échantillon de Paris**

Le texte total des 1000 réponses contient 11108 occurrences de 1229 formes distinctes (dont le nombre représente donc 11,1% du texte).

Le tableau 4 montre les 112 formes qui apparaissent au moins 18 fois dans



l'échantillon des 1000 réponses aux deux questions ouvertes. Ces 112 formes donnent lieu à 7806 occurrences, soit 70% du texte total. Comme on pourra en juger à partir des échantillons de réponses caractéristiques, les réponses sont plus riches et plus diversifiées, et surtout plus longues, ce qui a amené à choisir un seuil de sélection des formes beaucoup plus élevé, dans la première phase

num	Mot	Employé	frq						
				37	ENTRECOTE	29	75	PAELLA	32
				38	ENTREE	31	76	PAS	21
1	A	197		39	ESCALOPE	34	77	PATES	60
2	AGNEAU	71		40	ET	53	78	PIZZA	33
3	AU	202		41	FARCIES	19	79	PLAT	38
4	AUX	137		42	FILET	33	80	PLATEAU	20
5	AVEC	117		43	FOIE	101	81	PLATS	80
6	BLANC	32		44	FONDUE	21	82	POISSON	133
7	BLANQUETTE	21		45	FRAIS	20	83	POISSONS	73
8	BOEUF	156		46	Frites	163	84	POMMES	161
9	BOURGUIGNON	44		47	FROMAGE	318	85	PORC	41
10	CAFE	33		48	FROMAGES	50	86	POULET	175
11	CANARD	71		49	FRUIT	33	87	PUREE	33
12	CASSOULET	22		50	FRUITS	105	88	RAGOUT	19
13	CAVIAR	25		51	FUME	59	89	REPAS	44
14	CHAMPIGNONS	26		52	GATEAU	60	90	RIZ	112
15	CHARCUTERIE	59		53	GIGOT	50	91	ROTI	80
16	CHOCOLAT	72		54	GLACE	62	92	ROUGE	22
17	CHOUCRROUTE	32		55	GRAS	96	93	SALADE	357
18	CITRON	19		56	GRATIN	30	94	SALADES	35
19	COMPOSEE	22		57	GRILLE	36	95	SAUCE	80
20	COQUILLES	23		58	GRILLEE	27	96	SAUMON	143
21	COTE	35		59	HARICOTS	61	97	SOLE	29
22	COTES	31		60	HOMARD	36	98	SOUPE	40
23	COUSCOUS	50		61	HUITRES	48	99	SPAGHETTI	22
24	CREME	64		62	JACQUES	30	100	ST (Saint)	35
25	CREVETTES	31		63	L	62	101	STEACK	43
26	CRUDITES	94		64	LA	129	102	STEAK	145
27	CRUSTACES	25		65	LAPIN	29	103	TARTE	89
28	CUISINE	23		66	LEGUMES	155	104	TERRE	107
29	D	85		67	LES	24	105	TOMATES	27
30	DAUPHINOIS	19		68	MER	37	106	TOUS	20
31	DE	685		69	MEUNIERE	24	107	VEAU	59
32	DES	25		70	MOUSSE	29	108	VERTE	50
33	DESSERT	168		71	NeSaitPas	27	109	VERTS	65
34	DESSERTS	30		72	OEUFS	22	110	VIANDE	114
35	DU	22		73	OMELETTE	20	111	VIANDES	82
36	EN	74		74	OU	38	112	VIN	44

**Tableau 4 : PARIS : formes graphiques par ordre alphabétique ; seuil = 18**

d'exploitation présentée ici.

On note dans cette liste les fréquences importantes de mots outil et des termes génériques (bœuf, fromage, salade, viande). Malgré la sévérité du seuil, cette liste suffit pour reconstituer avec régularité les dispositions des catégories de la variable composite sexe-âge.

La figure 2 représente le plan des axes (1, 2) issu du tableau de contingence croisant les six catégories de cette variable avec les 112 formes.

La régularité de la structure mise en évidence tranche avec la forme trapézoïdale marquée de la figure 1, comme si les écarts entre sexes selon les générations étaient beaucoup moins sensibles à Paris qu'à Tokyo. Ces

	LIBELLE DE LA FORME GRAPHIQUE	POURCENTAGE INTERNE GLOBAL	FREQUENCE INTERNE GLOBALE	V. TEST	PROBA		
<i>moins de 30 ans - HOMMES</i>							
1	FRITES	4.04	2.09	50.	163.	4.704	0.000
2	PIZZA	1.13	0.42	14.	33.	3.486	0.000
3	GRAS	2.26	1.23	28.	96.	3.205	0.001
4	FOIE	2.26	1.29	28.	101.	2.948	0.002
5	CREME	1.54	0.82	19.	64.	2.671	0.004
6	ESCALOPE	0.89	0.44	11.	34.	2.231	0.013
<i>de 30 à 50 ans - HOMMES</i>							
1	GRATIN	0.80	0.38	9.	30.	2.007	0.022
2	DESSERT	2.92	2.15	33.	168.	1.760	0.039
3	STEAK	2.56	1.86	29.	145.	1.735	0.041
4	DAUPHINOIS	0.53	0.24	6.	19.	1.683	0.046
5	ENTRECOTE	0.71	0.37	8.	29.	1.655	0.049
6	FRITES	2.74	2.09	31.	163.	1.518	0.065
<i>plus de 50 ans - HOMMES</i>							
1	GRILLE	1.08	0.46	13.	36.	2.910	0.002
2	CASSOULET	0.75	0.28	9.	22.	2.702	0.003
3	DE	10.76	8.78	129.	685.	2.529	0.006
4	BOEUF	2.92	2.00	35.	156.	2.270	0.012
5	PLAT	0.92	0.49	11.	38.	1.977	0.024
6	PLATEAU	0.58	0.26	7.	20.	1.967	0.025
<i>moins de 30 ans - FEMMES</i>							
1	BLANC	0.88	0.41	11.	32.	2.375	0.009
2	STEACK	1.04	0.55	13.	43.	2.176	0.015
3	SPAGHETTI	0.64	0.28	8.	22.	2.120	0.017
4	RIZ	2.15	1.43	27.	112.	2.108	0.018
5	CREVETTES	0.80	0.40	10.	31.	2.063	0.020
6	FONDUE	0.48	0.27	6.	21.	1.237	0.108
<i>de 30 à 50 ans - FEMMES</i>							
1	PLAT	1.09	0.49	15.	38.	3.019	0.001
2	RIZ	2.17	1.43	30.	112.	2.314	0.010
3	SALADE	5.79	4.57	80.	357.	2.266	0.012
4	OU	0.87	0.49	12.	38.	1.932	0.027
5	COUSCOUS	1.01	0.64	14.	50.	1.670	0.047
6	POULET	2.82	2.24	39.	175.	1.488	0.068
<i>plus de 50 ans - FEMMES</i>							
1	SOUPE	1.44	0.51	23.	40.	4.987	0.000
2	NSP	0.87	0.35	14.	27.	3.434	0.000
3	LEGUMES	2.93	1.99	47.	155.	2.832	0.002
4	BOURGUIGNON	1.00	0.56	16.	44.	2.293	0.011
5	LAPIN	0.69	0.37	11.	29.	1.987	0.023
6	AGNEAU	1.37	0.91	22.	71.	1.973	0.024

**Tableau 5: Choix des formes caractéristiques (PARIS)**

différences accentuées entre générations ont déjà été vues lors d'autres enquêtes au Japon, mais il reste à vérifier la réalité des différences de structures observées.

Les tableaux 5 et 6 donnent les formes, puis les réponses caractéristiques de chaque catégorie.

Comme à Tokyo, la cuisine étrangère intéresse plutôt les jeunes. Autres points de convergence avec Tokyo, l'attrait des jeunes femmes pour la cuisine

*moins de 30 ans - HOMMES*

- 1 STEACK FRITES, POULET FRITES, SOLE MEUNIERE
- 1 FOIE GRAS TRUFFE, SAUMON A L'OSEILLE, FROMAGE, GLACE
- 2 PATES BOLOGNAISES, POULET FRITES, STEACK, SALADE
- 2 CAROTTES RAPEES, FOIE GRAS, DINDE MARRONS, POMMES FRITES, SALADE VERTE,
- 2 ECHALOTTES, GLACE

*de 30 à 50 ans - HOMMES*

- 1 STEAK, SALADE, FROMAGE, DESSERT, ENTRECOTE GRILLEE A LA SAUCE AU POIVRE VERT
- 1 AVEC POMMES DE TERRE, FROMAGE, CREME, TARTE
- 2 ENTRECOTE DE BOEUF, GRATIN DAUPHINOIS, RELIGIEUSE AU CHOCOLAT
- 2 PIEMONTAISE, HARENG FUME, POMMES A L'HUILE, SALADE VERTE, FROMAGE DOUX

*plus de 50 ans - HOMMES*

- 1 LEGUMES, VIANDE, PAVE DE BOEUF GRILLE AUX HERBES, SALADE DE GESIERS DE CANARD
- 2 CIVET DE LAPIN, LEGUMES, PLAT DE FRUITS DE MER, JAMBON DE PARME, CASSOULET,
- 2 FROMAGE, SALADE, MOUSSE AU CHOCOLAT

*moins de 30 ans - FEMMES*

- 1 PATES A LA TOMATE, ROTI DE VEAU, SAUMON
- 1 FOIE GRAS, SAUMON AU BLANC DE POIREAU, SAUTERNE, STEAK AU POIVRE VERT,
- 1 BAVOISE FRAMBOISE, BORDEAUX
- 2 RIZ SAUCE GOMBO, FRITE, STEACK, TOMATE FARCIE
- 2 AVOCAT A LA VINAIGRETTE, DINDE AU MARRON, GLACE

*de 30 à 50 ans - FEMMES*

- 1 PLAT EN SAUCE, PALETTE, CUISSE DE DINDE, COQUILLES ST JACQUES A LA PROVENCEALE,
- 1 FILETS MIGNONS AUX MORILLES, SALADE, ILES FLOTTANTES
- 2 PLAT EXOTIQUE
- 2 SAUMON FUME, CONFIT DE CANARD, SALADE, CHARLOTTE AUX NOISETTES

*plus de 50 ans - FEMMES*

- 1 BOEUF, SOUPE DE LEGUMES, COTES D'AGNEAU, POISSONS
- 1 CRUDITES, VIANDE DE POISSON, LEGUMES, SALADE, DESSERT
- 2 SOUPE DE LEGUMES, BOEUF BOURGUIGNON, SALADE, SAUMON FUME SUR TAOST,
- 2 TRUITE SAUMONNEE POMMES DE TERRE, FROMAGE, TARTE AUX MYRTILLES

**Tableau 6 : individus ou réponses caractéristiques (PARIS)**

italienne , et les mentions par les personnes plus âgées de plats nationaux traditionnels, (ici représentés par le cassoulet, le bœuf bourguignon).

Les régimes sont traditionnellement plus riches en viande en France qu'au Japon, mais on observe que, dans les deux pays, les femmes sont plutôt moins "carnivore" que les hommes, et les personnes âgées plutôt moins que les jeunes. Cette constatation est cependant troublée par deux facteurs (liés): les repas de fête, qui contiennent de la viande en France, et la préparation de repas pour la famille : le repas préféré est aussi celui que l'on aime préparer pour les autres. Peut-être la mention exceptionnelle de viandes par les femmes âgées en France relève-t-elle de cette confusion entre diététique personnelle et repas familial. Ceci reste à vérifier. On note l'attachement des jeunes aux menus traditionnels des repas de réveillon, tradition qui n'a pas, semble-t-il, d'équivalent ou d'analogue dans les deux autres villes enquêtées. Les produits de restauration rapide à l'américaine ne sont pas absents du corpus, mais n'apparaissent pas au niveau des formes et des réponses les plus caractéristiques, ce qui constitue une différence importante avec Tokyo, et, on va le voir, avec New-York.

num	Mot	Employé	frq						
1	AND	176	28	FRENCH	41	56	POTATO	151	
2	APPETIZER	17	29	FRESH	22	57	POTATOES	112	
3	APPLE	23	30	FRIED	75	58	RED	17	
4	BAKED	131	31	FRIES	40	59	RIBS	14	
5	BEANS	59	32	FRUIT	38	60	RICE	118	
6	BEEF	92	33	GARLIC	30	61	ROAST	89	
7	BREAD	63	34	GRAVY	20	62	SALAD	348	
8	BROCCOLI	36	35	GREEN	34	63	SALADS	13	
9	BROILED	55	36	HAMBURGER	28	64	SALMON	22	
10	CAESAR	16	37	HAMBURGERS	38	65	SANDWICH	15	
11	CAKE	42	38	HOT	18	66	SAUCE	39	
12	CHEESE	33	39	ICE	37	67	SEAFOOD	29	
13	CHEESECAKE	15	40	ITALIAN	65	68	SHRIMP	109	
14	CHICKEN	393	41	LAMB	16	69	SOUP	107	
15	CHINESE	72	42	LASAGNA	22	70	SPAGHETTI	87	
16	CHOCOLATE	34	43	LOBSTER	100	71	STEAK	207	
17	CHOPS	27	44	MACARONI	55	72	STEAMED	16	
18	CLAM	16	45	MASHED	51	73	STEW	23	
19	CLAMS	19	46	MEAT	52	74	STUFFED	15	
20	COCKTAIL	28	47	MEATBALLS	40	75	TUNA	36	
21	COFFEE	47	48	ONION	16	76	TURKEY	37	
22	CORN	19	49	ONIONS	13	77	VEAL	65	
23	CREAM	44	50	PARMESAN	52	78	VEGETABLE	70	
24	CUTLETS	13	51	PASTA	129	79	VEGETABLES	123	
25	DESSERT	47	52	PIE	34	80	WHITE	13	
26	FISH	145	53	PIZZA	64	81	WINE	33	
27	FOOD	117	54	PORK	27	82	WITH	159	
			55	POT	15	83	ZITI	16	

Tableau 7: New-York: formes graphiques par ordre alphabétique ; seuil = 12

### 3 Sous-échantillon de New-York

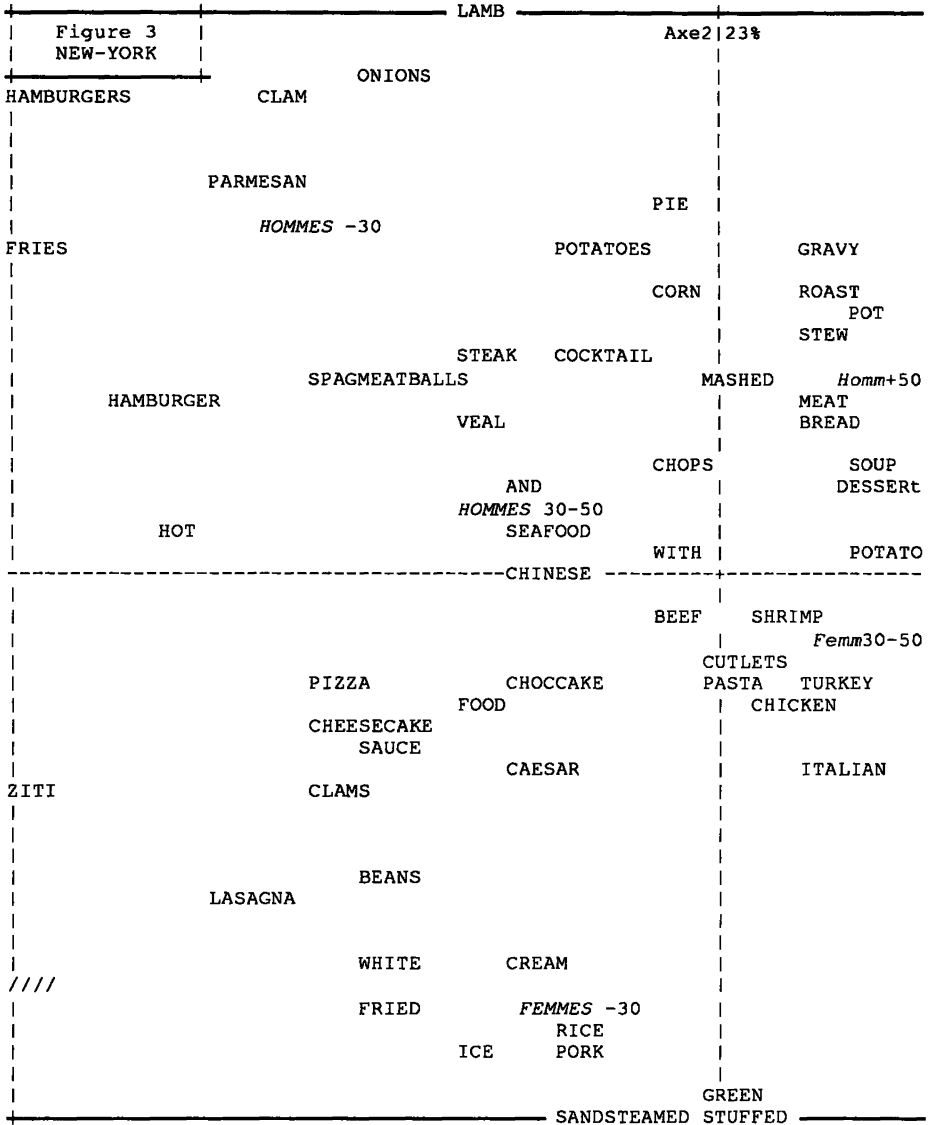
Le sous-échantillon des personnes interrogées à New-York est sensiblement plus réduit que les précédents (634 personnes).

Le corpus des réponses libres contient 6511 occurrences de 638 formes distinctes. Sur ces 6511 occurrences, 5034 (soit 77%) concernent les 83 formes les plus fréquentes (apparaissant plus de 12 fois), qui serviront de base à notre étude (tableau 7).

La figure 3, qui reproduit, comme les figures 1 et 2, le plan (1, 2) issu de l'analyse du tableau de contingence croisant formes et catégories montre une configuration de points plus tourmentée que celles de Paris et Tokyo, mais cependant très similaire, si l'on excepte la position des jeunes hommes, "trop" hauts sur la partie gauche du graphique.

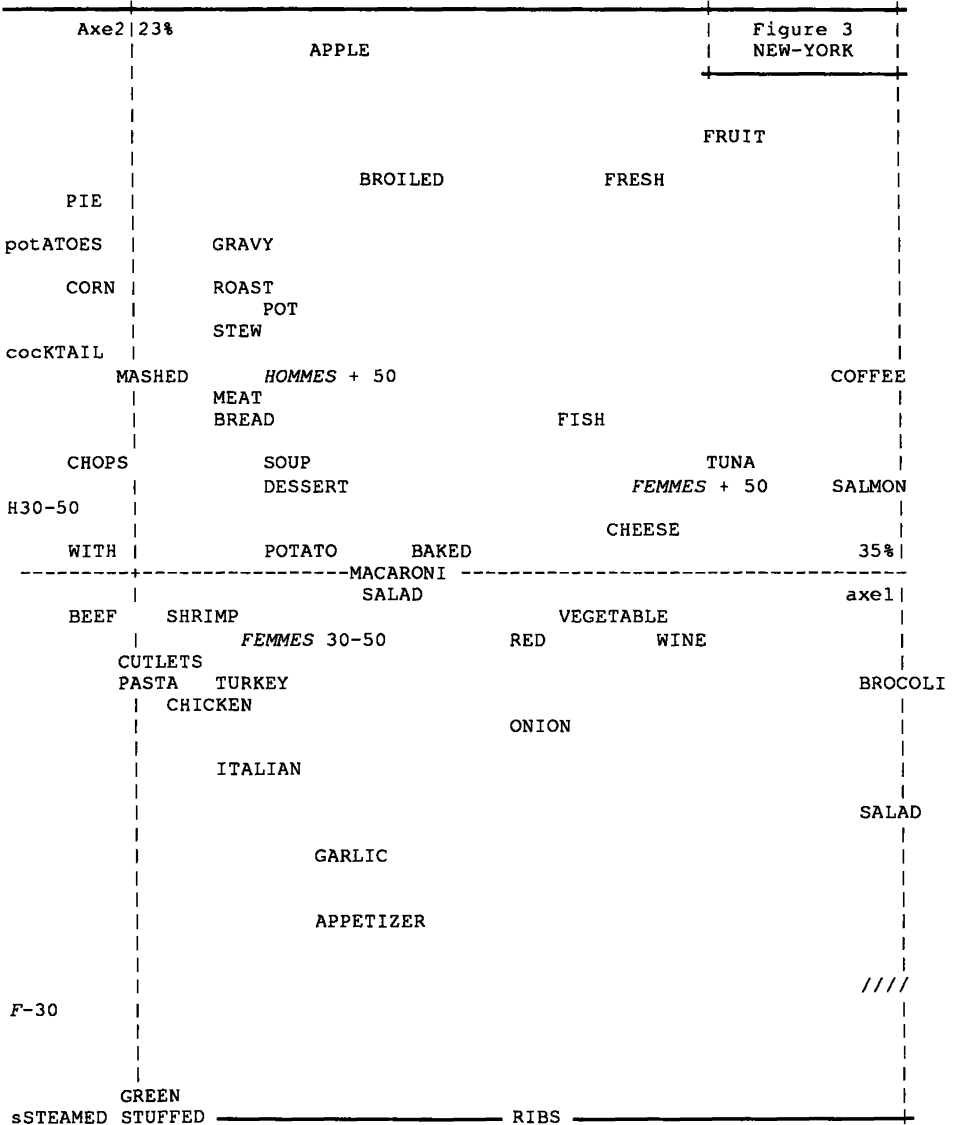
Un peu comme à Tokyo, les différences semblent fortes entre les hommes et les femmes de moins de 30 ans. Dans les deux villes, l'attrance des jeunes garçons pour les "fast foods" est une cause importante de divergence (l'absence





relative de différences entre les garçons et les filles à Paris pourrait provenir de la faible importance de ce type de repas).

Notons que contrairement à ce qui est le cas dans les figures précédentes, ce n'est plus le premier axe qui sépare les hommes des femmes, mais la diagonale



principale.

La sélection des formes caractéristiques (tableau 8) confirme assez largement le sens des proximités observées sur la figure 3, avec quelques corrections de l'approximation plane : la forme "CHICKEN" (qui est la forme la

LIBELLE DE LA FORME GRAPHIQUE	POURCENTAGE		FREQUENCE		V. TEST	PROBA
	INTERNE	GLOBAL	INTERNE	GLOBALE		
<i>moins de 30 ans - HOMMES</i>						
1 HAMBURGERS	2.76	0.75	14.	38.	4.284	0.000
2 POTATOES	4.34	2.22	22.	112.	2.968	0.001
3 PARMESAN	2.37	1.03	12.	52.	2.606	0.005
4 FRIES	1.97	0.79	10.	40.	2.566	0.005
5 PIZZA	2.37	1.27	12.	64.	1.977	0.024
6 MASHED	1.97	1.01	10.	51.	1.904	0.028
<i>de 30 à 50 ans - HOMMES</i>						
1 SPAGHETTI	2.84	1.73	31.	87.	2.896	0.002
2 VEAL	2.20	1.29	24.	65.	2.696	0.004
3 COCKTAIL	1.10	0.56	12.	28.	2.340	0.010
4 FOOD	3.29	2.32	36.	117.	2.219	0.013
5 LOBSTER	2.84	1.99	31.	100.	2.087	0.018
6 STEAK	5.22	4.11	57.	207.	1.951	0.026
<i>plus de 50 ans - HOMMES</i>						
1 SOUP	3.84	2.13	29.	107.	3.167	0.001
2 CORN	0.93	0.38	7.	19.	2.141	0.016
3 COFFEE	1.59	0.93	12.	47.	1.745	0.040
4 POTATOES	3.18	2.22	24.	112.	1.740	0.041
5 STEW	0.93	0.46	7.	23.	1.686	0.046
6 FRUIT	1.32	0.75	10.	38.	1.657	0.049
<i>moins de 30 ans - FEMMES</i>						
1 RICE	4.61	2.34	33.	118.	3.835	0.000
2 FRIED	2.93	1.49	21.	75.	3.009	0.001
3 CHICKEN	10.06	7.81	72.	393.	2.287	0.011
4 SANDWICH	0.84	0.30	6.	15.	2.228	0.013
5 GREEN	1.40	0.68	10.	34.	2.128	0.017
6 STEAMED	0.84	0.32	6.	16.	2.091	0.018
<i>de 30 à 50 ans - FEMMES</i>						
1 PASTA	3.33	2.56	38.	129.	1.727	0.042
2 VEGETABLE	1.93	1.39	22.	70.	1.586	0.056
3 CHOCOLATE	1.05	0.68	12.	34.	1.519	0.064
4 DESSERT	1.31	0.93	15.	47.	1.327	0.092
5 FISH	3.51	2.88	40.	145.	1.324	0.093
6 SEAFOOD	0.88	0.58	10.	29.	1.282	0.100
<i>plus de 50 ans - FEMMES</i>						
1 SALAD	10.34	6.91	85.	348.	3.967	0.000
2 BROCCOLI	1.70	0.72	14.	36.	3.105	0.001
3 COFFEE	1.95	0.93	16.	47.	2.846	0.002
4 TUNA	1.58	0.72	13.	36.	2.734	0.003
5 VEGETABLE	2.43	1.39	20.	70.	2.471	0.007
6 FISH	4.14	2.88	34.	145.	2.159	0.015

Tableau 8: Choix des formes caractéristiques (NEW-YORK)

plus fréquente de tout le corpus, avec 393 occurrences), par exemple, caractérise plus les femmes de moins de 30 ans que ne le laisse deviner sa position assez centrale sur le graphique.

Notons également que les deux catégories ayant les valeurs-test les plus fortes sont aussi les deux catégories extrêmes sur le premier facteur : les hommes de moins de 30 ans et les femmes de plus de 50 ans.

Le tableau 9 confirme, au niveau des réponses dans leur ensemble, l'homogénéité interne et la dissemblance des classes extrêmes.

*moins de 30 ans - HOMMES*

- 1 HAMBURGERS, PIZZA, SPAGHETTI
- 1 SALAD, BROILED STEAK, FRENCH
- 1 FRIES, COLE SLAW
- 2 HAMBURGERS, PIZZA, SPAGHETTI
- 2 STEAK, FRENCH FRIES
- 3 PIZZA, HAMBURGERS, FRIES
- 3 SPAGHETTI WITH MEATBALLS, SALAD, CAKE, COFFEE

*de 30 à 50 ans - HOMMES*

- 1 VEAL MARSALA, CHINESE FOOD, SHRIMP IN LOBSTER SAUCE
- 1 ESCARGOT, VEAL, BAKED POTATO, CHOCOLATE MOUSSE
- 2 ITALIAN FOOD, VEAL AND PEPPERS, SPAGHETTI AND MEATBALLS
- 2 VEAL AND PEPPERS, BREAD, CHOCOLATE CAKE
- 3 STEAK, SHRIMP PARMESAN, LOBSTER CANTONESE
- 3 LOBSTER CANTONESE, EGG DROP SOUP, FRIED RICE, PEPPER STEAK

*plus de 50 ans - HOMMES*

- 1 STEAK, CHICKEN, FISH BAKED CHICKEN, VEGETABLES, POTATOES, SALAD
- 2 ROAST CHICKEN, VEGETABLES, FISH, SOUP, MEAT, VEGETABLES, SALAD
- 3 BROILED CHICKEN, FRESH FISH, ROAST BEEF
- 3 FRUIT CUP, ROAST , CHICKEN WITH BAKED POTATO, SALAD, SHERBET

*moins de 30 ans - FEMMES*

- 1 FRIED CHICKEN, CHICKEN TERIYAKI, CHICKEN SALAD
- 1 FRIED CHICKEN, STEAMED VEGETABLES, CHOCOLATE CAKE
- 2 RICE AND BEANS, GREEN BANANAS, CHICKEN AND RICE
- 2 RICE AND BEANS, GREEN SALAD, BEER, CHICKEN AND RICE
- 3 FRIED CHICKEN, STEAKS, RICE AND BEANS, FRIED SHRIMP, DESSERT

*30 à 50 ans - FEMMES*

- 1 CHICKEN, FISH, STEAK, WINE, SALAD, SURF AND TURF, BAKED POTATO
- 2 CHICKEN, ITALIAN FOOD, ROAST BEEF AND BAKED POTATO
- 2 GREEN SALAD, BAKED CHICKEN, BROCCOLI, ICE CREAM
- 3 CHICKEN, POT ROAST, BAKED FISH
- 3 LOBSTER BISQUE, SALAD, LOBSTER, BAKED POTATO

*plus de 50 ans - FEMMES*

- 1 BAKED SALMON, TUNA SALAD, FILET OF SOLE
- 1 FISH, BAKED POTATO, VEGETABLE, FRESH FRUIT JELLO
- 2 TUNA SALAD, PIZZA, BBQ CHICKEN
- 2 MATZAH BALL SOUP, BBQ CHICKEN, BAKED POTATO, BROCCOLI
- 3 COLE SLAW, CABBAGE SALAD, CARROT RAISIN SALAD FRUIT AND
- 3 VEGETABLE SALAD, BREAD, COFFEE, WATERMELON

**Tableau 9: individus ou réponses caractéristiques (NEW-YORK)****4. Discussion**

Ces résultats constituent une première étape de travail, qui encourage cependant à approfondir les traitements et analyses. Mais dès cette première étape, les faits statistiques mis en évidence ainsi que leur interprétation posent d'emblée un certain nombre de problèmes que nous allons exposer et discuter en suivant un canevas inspiré directement par trois remarques de J.-P. Benzécri, que nous résumons dans les termes suivants:

a) Peut-on, dans le cas du Japon, affirmer que les personnes âgées des deux sexes sont bien plus proches entre elles que les classes d'âges plus jeunes (le troisième facteur de la figure 1 pourrait en effet transformer le triangle en tétraèdre, et éloigner ces deux points pourtant proches en projection) ?

Plus généralement, quelle est la réalité des configurations observées, et quelles sont les limites des interprétations entre différences de configurations ?

b) Que donne une analyse directe du tableau non-agrégé "réponses-mots" (ou ce qui revient au même, pour les proximités entre mots, de la matrice de cooccurrences entre mots) ? Sur une telle analyse, on peut en effet projeter en éléments supplémentaires les catégories des diverses variables nominales décrivant les individus. Peut-on espérer alors retrouver les configurations des figures 1 à 3 ?

c) Des classifications hiérarchiques opérées sur les formes graphiques seraient un utile complément aux traitements précédents, en mettant probablement en évidence plusieurs niveaux d'associations que le traitement de tableaux agrégés ne permet pas d'observer.

En fait, les points b et c qui concernent les associations entre mots à l'intérieur des réponses individuelles seront traités ensemble, car ils s'éclairent mutuellement.

#### 4.1 Réalité des configurations (point a)

La comparaison des configurations des points "âge-sexe" des figures 1 à 3 est évidemment hypothéquée par la qualité de la projection de l'espace à 5 dimensions initial dans le plan factoriel. La meilleure façon de lever cette hypothèque nous a paru être de décrire par classification hiérarchique les distances entre points catégories, en utilisant le critère d'agrégation suivant la variance, (dit de Ward).

La figure 4 (qu'on a placée à la fin du §1, afin d'éviter que les graphiques du §2 ne se mêlent au texte du §1) montre les trois dendrogrammes de Tokyo, Paris et New-York. Les indices correspondant à chaque noeud sont exprimés en pourcentages de la somme des indices.

On note par exemple que, pour Tokyo, la "convergence des sexes avec l'âge" constatée sur la figure 1 (convergence responsable de la forme trapézoïdale de la configuration) est bien visible sur la figure 4: les plus âgés s'agrègent pour un niveau bas de l'indice (11% de la somme), et les plus jeunes à un niveau plus élevé (22.5%). C'est effectivement à Tokyo que les personnes les plus âgées ont les comportements les plus homogènes (niveau 11%, contre 16% à New-York et 19% à Paris).

Une autre particularité de Tokyo concerne le rattachement tardif des jeunes au reste de la population : la coupure la plus nette est à 30 ans dans cette ville, alors qu'elle est à 50 ans dans les deux autres. A Paris et New-York, le rattachement des personnes les plus âgées au reste de la population se fait au plus haut niveau du dendrogramme.

On remarque que les femmes d'une classe d'âge ne se rattachent jamais à des hommes d'une classe d'âge inférieure : à Tokyo, le point F-50 rejoint les points H+50 et F+50 avant que ne le fasse le point H-50; à Paris, le point F-30 s'agrège très bas au point H-50; à New-York, le point F-30 s'agrège à la classe {F-50, H-50} avant de s'agréger au point H-30.

On peut d'ailleurs vérifier sur les figures 1 à 3 que sur les axes d'âge, les femmes sont systématiquement décalées par rapport aux hommes dans le sens d'un âge plus élevé. (décalage qui a sa contrepartie dans les préférences alimentaires, puisque les femmes semblent avoir plus jeunes des goûts pour les légumes, les salades et les poissons).

L'originalité des jeunes hommes à New-York, déjà visible sur la figure 3, est confirmée par le dendrogramme. L'analogie avec Tokyo est plus faible que ce que l'on pouvait imaginer au vu des seules figures 1 et 3.

Les classifications ont donc permis, selon les cas, de confirmer ou de nuancer les premières assertions interprétatives.

#### **4.2 Analyses directes des réponses individuelles (points b, c)**

Nous illustrerons ce paragraphe, de caractère méthodologique, par le volet français de l'enquête.

La relative richesse des réponses nous suggère de travailler avec un seuil de fréquence des formes plus faible : 10 (au lieu de 18 au §2) ce qui conduit à retenir 173 formes correspondant à un ensemble 8632 occurrences (au lieu de 7806, au §2).

##### **4.2.1 Analyse de la correspondance entre individus et mots**

L'analyse des correspondances du tableau de contingence T décrivant les réponses individuelles (ici tableau à 1000 lignes et 173 colonnes), en adjoignant les variables nominales disponibles en éléments supplémentaires, constitue en général, pour nous, la première approche systématique des données de ce type. C'est ainsi que nous sélectionnons les catégories les plus pertinentes, catégories utilisées par la suite pour regroupements des réponses, comme on l'a fait ici pour la variable composite sexe-âge. Les modalités de cette variable occupent en effet des positions significatives sur certains des dix premiers axes.

Ce choix ne peut se faire uniquement sur les tout premiers axes issus de l'analyse de T, instables, compte tenu du caractère clairsemé du tableau de données (moins de 5% de cases non-nulles pour le tableau  $1000 \times 173$ ). Ainsi, pour le tableau T, la première valeur propre (0.4703) explique 2.22% de la trace (21.16), les 15 premières en expliquant seulement 22.6%. La différence entre les valeurs propres 30 (0.1989) et 29 (0.1957) est de 0.0032. L'analyse de ce tableau est sensiblement équivalente à celle du tableau ( $173 \times 173$ ) décrivant les cooccurrences de mots à l'intérieur des réponses.

1| A CREME ESCALOPE LA MOULES PROVENCALE THON TOMATE  
 2| ARMORICAINE CANARD CONFIT L MAGRET ORANGE  
 3| CAVIAR CHARLOTTE CHEVREUIL DINDE FOIE FUME GLACES GRAS HOMARD HUITRES  
 | SAUMON  
 4| GIBIER MEUNIERE SOLE  
 5| FONDUE FORET NOIRE  
 6| LASAGNES SPAGHETTI  
 7| FRITES HACHE PATES POIVRE SORBET STEACK STEAK  
 8| FARCIES JAMBON OEUFS TOMATES  
 9| BLANQUETTE MOUTON RAGOUT VEAU  
 10| FRUITS MER PLATEAU TARTES  
 11| AUX CHEVRE OMELETTE PATISSERIE TARTE TATIN  
 12| DE ESCARGOTS LANGOUSTE POIS POMME POMMES PORC PUREE SAUTEES TERRE  
 13| FRUIT GRILLEE HARICOTS LEGUMES POTAGE VERTS  
 14| CAFE CAROTTES COMPOSEE FROMAGE LEGUME PETITS POULET ROTI SALADE VERT  
 | VERTE  
 15| APERITIF CHAMPIGNONS CHOUX COTE EN ENTRECOTE FILET LAPIN PATE PAVE  
 | SAUCE TERRINE  
 16| AVEC DES DU ET UN  
 17| BLANC BOURGOGNE CREVETTES CURRY FRAIS GLACE GRILLE OU POISSON RIZ  
 | SOUPE TRUITE VAPEUR VIN  
 18| BOEUF BOURGUIGNON CRU  
 19| AU BEURRE CHOCOLAT CITRON DAUPHINOIS ENDIVES FOUR GATEAU GRATIN MOUSSE  
 20| AGNEAU COTES D FLAGEOLETS GIGOT  
 21| CREPES PIZZAS  
 22| PIZZA  
 23| COQUILLES JACQUES ST (Saint)  
 24| COUSCOUS  
 25| CHOUCROUTE  
 26| LEGER  
 27| CASSOULET CRUDITES FEU GRILLADES POT SALADES  
 28| CRUSTACES PAELLA POISSONS  
 29| DESSERT ROUGE VIANDE  
 30| ENTREE PLAT  
 31| CHARCUTERIE PARMENTIER  
 32| ENTREES  
 33| CRUDITE DESSERTS FROMAGES VIANDES  
 34| PAS  
 35| LES TOUS  
 36| EQUILIBRE REPAS  
 37| PLATS  
 38| CUISINE  
 39| SPECIALITES  
 40| NSP (ne sait pas)

**Tableau 10: Partition en 40 classes, (PARIS)**

#### 4.2.2 Classification des mots

Une classification hiérarchique des mots illustrera le fait que l'essentiel de l'inertie inter-réponses n'est pas toujours pertinent pour comparer des agrégats importants d'individus. Les premiers facteurs de l'analyse directe saisissent souvent une variabilité entre réponses qui est une variabilité intra-catégories (par opposition à la variabilité inter-catégories qui nous a principalement intéressés au cours des pages précédentes). Les associations entre mots doivent en effet beaucoup aux règles syntaxiques, à des locutions ou associations figées; ou, dans le cas de notre corpus, à des plats composés ou des recettes classiques...

Le tableau 10 présente une partition en 40 classes, issue de celle définie par les 39 nœuds les plus hauts de la CAH, mais élaborée après un certain nombre d'itérations par agrégation autour de centres mobiles, qui ont pour effet de réaffecter quelques formes ou groupes de formes.

Au bas du tableau 10, se trouvent des classes qui se séparent à un niveau très haut de la hiérarchie: il s'agit de branchements qui correspondent, en fait, aux premiers axes factoriels de l'analyse des correspondances du tableau T.

Ainsi le premier axe est caractérisé par la forme NSP (ne sait pas), qui est parfois une des deux réponses libres sollicitées, ou une réponse à une relance. Cette forme - particulière, il est vrai - est associée à des contextes très variés, tout comme les autres formes du haut de l'arbre qui correspondent aux axes suivants: CUISINE, SPECIALITES, (chinoises, italiennes, régionales, maison), PLATS, et comme les formes souvent associées TOUS, LES (tous les plats, tous les desserts, tous les fruits, tous les poissons) et REPAS, EQUILIBRE, ainsi que PAS (dans ne sait pas, pas de préférence, pas de repas idéal...). Des segments contraints (selon la terminologie de A.Salem), comme Coquille St Jacques, ont aussi des contributions majeures aux premiers axes, comme d'autres segments répétés plus libres (gigot d'agneau, côtes d'agneau).

En somme, l'analyse factorielle d'un tableau clairsemé, glanant quelques petits conglomerats à la surface d'une gigantesque hypersphère, est très liée aux niveaux élevés d'une classification hiérarchique. Mais les parties médianes et basses de l'arbre fournissent également une description très utile des associations entre mots. La classe 16, qui correspond aux terminaux de rangs 92 à 96 de l'arbre, est un exemple étonnant de squelette de réponse, formé uniquement de mots-outil. Le mot DE (le plus fréquent: 685 occurrences) va, avec les segments qui le contiennent le plus fréquemment (classe 12), comme AU avec la classe 19, AUX avec la classe 11, {A LA} avec la classe 1, etc...

Comme c'est le cas d'ordinaire, ce type de description met en relief les dangers d'une lemmatisation aveugle. On note en particulier que les pluriels et les singuliers ne sont pas souvent dans les mêmes classes.

### 4.2.3 Stabilité de la configuration

Bien que les axes d'une telle analyse ne soient pas indépendants des catégories "sexe-âge" précédentes, (par exemple: les femmes de plus de 50 ans répondent souvent NSP), il est improbable de voir se dessiner sur des dimensions somme toute anecdotiques des configurations régulières et stables.

Pour comparer les configurations, nous reviendrons au seuil 18 utilisé dans la section 2, c'est-à-dire à un vocabulaire de 112 formes. (Notons d'ailleurs que l'arbre obtenu par classification de ces 112 formes est extrêmement voisin de celui obtenu en supprimant les branches correspondant aux formes non retenues dans l'arbre des 173 formes).



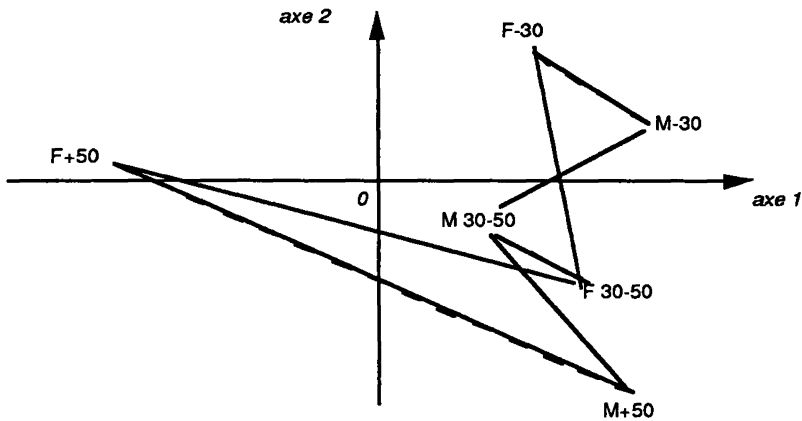


Figure 5: Analyse directe du tableau (1000 × 112): points "sexe-âge" dans le plan (1, 2)

La figure 5 montre la configuration obtenue dans l'analyse directe du tableau T0 (1000 × 112), avec projection des catégories en éléments supplémentaires: axe 1:  $\lambda = 0.482$ ,  $\tau = 3.14\%$ , axe 2:  $\lambda = 0.433$ ,  $\tau = 2.83\%$ . Le premier axe oppose la forme NSP (à gauche) à toutes les autres. Sur ce premier axe, la forme NSP a une contribution de 94% ; sur le second, on note CUISINE (50%), PLATS (14%), REPAS (17%), que l'on reconnaît comme caractéristiques des dernières branches (qui se séparent au plus haut) de la CAH.

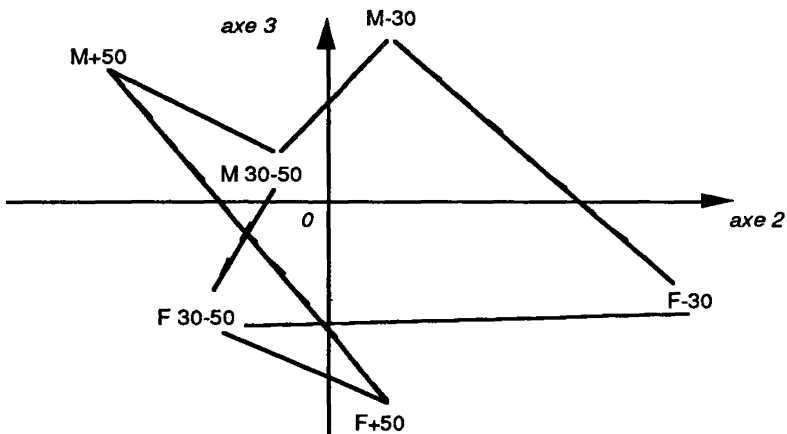


Figure 6: Analyse directe du tableau (1000 × 112): points "sexe-âge" dans le plan (2,3)

On peut penser que NSP est un élément de perturbation: la figure 6 montre que la situation s'améliore dans le plan (2,3). (axe 3 :  $\lambda = 0.373$ ,  $\tau = 2.43\%$ ).

Comme on pouvait le prévoir, une figure pratiquement identique à la figure 6 est obtenue par analyse du tableau (1000 × 111) restant après élimination de la forme NSP.

Enfin, si l'on supprime les formes qui constituent les classes 34 à 40 du tableau 10, (PAS, LÈS, TOUS, EQUILIBRE, REPAS, PLATS, CUISINE, SPECIALITES, NSP), on obtient un tableau (1000 × 105), (les formes *équilibré* et *spécialités* n'apparaissant pas avec le seuil 10).

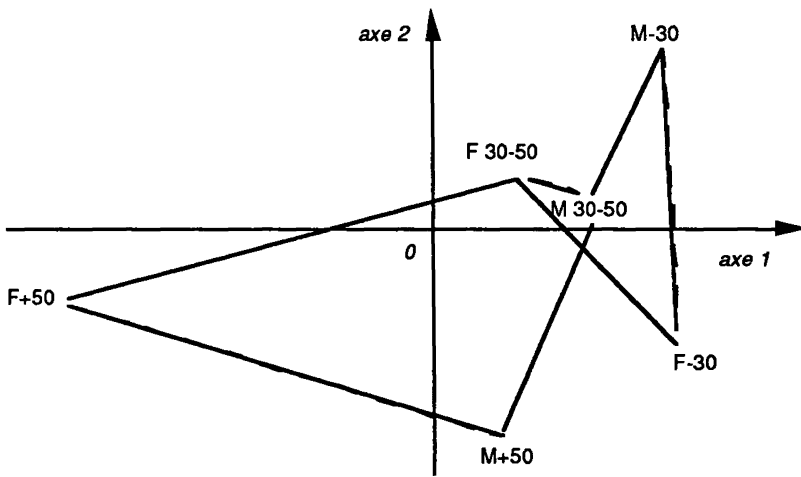


Figure 7: Analyse directe du tableau (1000 × 105): points "sexe-âge" dans le plan (1, 2)

L'analyse de ce tableau donne un plan (1, 2) représenté sur la figure 7.

(axe 1 :  $\lambda = 0.360$ ,  $\tau = 2.64\%$ , axe 2 :  $\lambda = 0.328$ ,  $\tau = 2.40\%$ )

Toutes ces analyses montrent l'instabilité de la configuration d'ensemble, mais aussi la permanence de certains traits structuraux : les classes extrêmes sont en général bien séparées.

Notons au contraire que la configuration de la figure 2, obtenue à partir du tableau agrégé, est remarquablement stable. Elle n'est pas affectée par la suppression des formes opérée plus haut.

L'analyse du tableau de contingence (6 × 105) produit des facteurs très proches de ceux issus de l'analyse du tableau initial (6 × 112). Il est clair que l'agrégation en catégories a pour effet d'effacer une partie importante de l'inertie inter-individu.

## 5 Conclusions et perspectives

Nos conclusions ne peuvent être que provisoires, car l'étude se poursuit.

La possibilité de comparer (assez facilement, au moins dans une première approche) des comportements à partir de textes dans des langues différentes ouvre la voie à d'intéressantes recherches. On esquiverait ainsi les phases de traduction, puis de codage, avec toute la part d'arbitraire, d'interprétation, d'erreurs qu'elles comportent; en fait, on n'aurait su que faire de traductions, car les variables de base ne sont pas les mêmes.

Les méthodes d'analyse des données ont montré, entre autres choses, que la multiplicité des variables permettait de travailler dans des espaces riches de sens, ou encore dans des espaces permettant des comparaisons plus fines que ne le permettrait chacune des variables initiales. Dans le type d'application présenté ici, on prend encore plus de recul par rapport aux variables initiales, dont on n'exige plus qu'elles soient identiques, ni même en nombres égaux.

Parce qu'on disposait de règles d'agrégation communes et d'échantillons relativement importants, on a ainsi pu procéder à des comparaisons sur un thème commun à partir de matériaux beaucoup plus hétérogènes que ceux issus des protocoles classiques. Les travaux complémentaires réalisés montrent la richesse de l'information de base, et l'étendue des difficultés que le chercheur doit s'attendre à rencontrer, même si, pour une langue donnée, ces difficultés sont recensées, sinon maîtrisables (choix des unités statistiques pertinentes, degré de lemmatisation, par exemple).

### Références bibliographiques

H. AKUTO (éd.): *International Comparison of Dietary Cultures*, Tokyo; (à paraître: 1992).

J. - P. BENZÉCRI & coll.: *Pratique de l'Analyse des Données en Linguistique et Lexicologie*, Dunod, Paris; (1981).

Ch. HAYASHI: "Theory and Examples of Quantification". (II) in *Proc. of the Institute of Stat. Math.*, 4 (2) 19-30; (1956).

L. LEBART, A. SALEM: *Analyse Statistique des Données Textuelles*, Dunod, Paris; (1988).

A. SALEM: *Pratique des Segments Répétés, Essai de Statistique Textuelle*, Klincksieck, Paris; (1987).